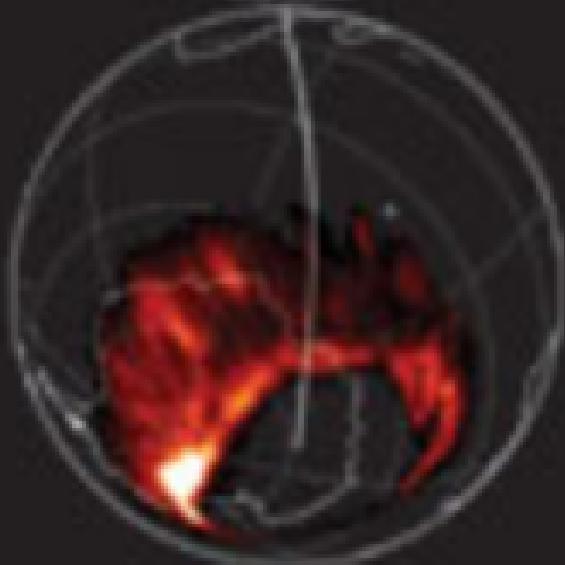


nature



DOUBLE EXPOSURE

Views from space show polar asymmetry of Earth's aurora



EUROPEAN EARTH

Earthquakes get real

INDIANO CANADA

Phenom at last seen

MONDAY

On schedule until August

See pages 466–467

See pages 468–469

Abstractions



LAST AUTHOR

In mammals, blood sugar levels must be carefully balanced: too low, and the brain and other organs are starved; too high, and nerves, blood vessels and organs may

be damaged. Thus, when mammals fast, even for brief periods such as during sleep, the pancreas releases glucagon, a hormone that stimulates the liver to produce glucose. Conversely, when mammals eat, an increase in pancreatic insulin instructs the liver to halt this activity. In obesity, this regulation tends to be lost, leading to insulin resistance — in which the liver becomes deaf to insulin's signals — and elevated blood sugar levels, or hyperglycaemia. Marc Montminy and his colleagues at the Salk Institute for Biological Studies in La Jolla, California, have now discovered a mechanism by which the protein CRTC2 (also known as TORC2) contributes to this problem (see page 534). Montminy tells *Nature* more.

What does CRTC2 do?

The CRTC2 protein functions as a genetic switch that triggers glucose production by the liver during fasting. We discovered that, in addition to responding to glucagon, CRTC2 can sense increases in stress in the endoplasmic reticulum (ER) — the cellular site that processes and folds newly synthesized proteins. Normally, ER stress acts as a brake on liver cells' production of glucose. But in obesity, the ER-stress pathway is chronically activated, so CRTC2's ability to cut glucose production is compromised. A component of the ER-stress brake, the protein ATF6, is also reduced, allowing the CRTC2 switch to boost glucose production, leading to hyperglycaemia and insulin resistance.

Can you fix this defect?

We were able to partly restore normal glucose regulation in obese mice by increasing the liver's production of ATF6. We inserted the ATF6 gene into an engineered adenovirus and injected this into obese mice. The virus then travelled to the liver, where it boosted the production of ATF6, restoring it to roughly normal levels. As a result, blood glucose levels stabilized.

How did you get the mice to become obese?

We fed them a Western-style diet in which 60% of the calories were provided by fat. After eight to ten weeks, the mice became obese and displayed physiological changes consistent with insulin resistance.

Does this work have implications for humans?

Developing drugs that modulate the ER-stress pathway in the liver could be of therapeutic benefit to obese individuals who are insulin resistant.

MAKING THE PAPER

Taylor Perron

Erosion theory explains uniform patterning of landscapes.

For more than a century, geologists have been aware that some landscapes bear seemingly uniform patterns, such as evenly spaced ridges and valleys. Taylor Perron, a geologist at the Massachusetts Institute of Technology in Cambridge, became intrigued by this ridge-valley 'wavelength' as he observed Earth's topography from aeroplanes, particularly over the central California Coast Ranges between San Francisco and Los Angeles. "Geologists," he says, "have a marked preference for window seats."

Perron and his colleagues, Jim Kirchner and Bill Dietrich at the University of California, Berkeley, decided to examine how erosional processes create these patterns by comparing a computational model with precise measurements of various landscapes. "We agreed that if we were going to try to explain how landscapes form, we'd have to be able to explain these patterns," says Perron.

The first step was to scout for sites — which they did with satellite images and Kirchner's Cessna aircraft — where they would be able to gain access on foot. "If you want to write the equations that describe how the topography evolves, it's important to stand there and see what's happening," says Perron. But accessing such sites required delicate negotiations with landowners, especially on agricultural and ranching lands, where soil erosion is a sensitive topic.

In the Salinas Valley of central California, where rows of lettuce, artichoke and other crops stretch as far as the eye can see, Perron and his team found a ranching family that granted them access to a section of Gabilan Mesa. This is a swath of land with relatively uniform valley spacing that covers thousands of square kilometres.

There, Perron and colleagues took a close-up look at the physical processes sculpting



the landscape, including gully erosion, which slowly deepens valleys by scouring soil, and 'bioturbation', such as ground-squirrel burrowing, which gradually smooths the landscape by stirring the soil and moving it downslope. "The basic mechanism is a competition between one process that incises valleys and another process that tends to smooth them out," says Perron.

This mechanism was also found to occur at four other sites the team examined, with differences in climate and rock type accounting for variations in wavelength from site to site.

Next, the authors turned to laser altimetry, a technique that relies on the Global Positioning System and aircraft-mounted lasers to produce precise digital topographic maps, even in landscapes covered by dense vegetation. Using this, the team precisely measured the ridge-valley wavelengths at the five field sites, and tested whether the measured wavelengths matched the predictions of a computational model built from their field observations. To Perron's excitement, they did (see page 502).

Having developed a computational model that accurately predicts the natural patterns in landscapes, geologists can learn more about how factors such as climate have shaped Earth's topography. And the model could potentially be applied farther afield — on Mars, for example, where evenly spaced ridges and valleys have formed on steep slopes such as those within impact craters. Knowing that similar topographical patterns occur on other planets, says Perron, "can tell us many things about those landscapes, even if we can't get there on the ground." ■

FROM THE BLOGOSPHERE

Neil Armstrong's "one small step" to the surface of the Moon has reverberated for the past 40 years. Freelance reporter

Lucas Laursen looks back at the Apollo 11 mission, twittering events in 'real time' as they happened in 1969 (<http://twitter.com/ApolloPlus40>).

In the accompanying blog series, @ApolloPlus40 (<http://tinyurl.com/lxp6vu>), Laursen has gathered stories of public

and political reactions to the missions, as well as the medical and technological doors it opened.

In one post, he covers a 14 July editorial that responded to then NASA administrator Thomas Paine's view that the mission was "a triumph of the squares" (a synonym for nerd) by promoting the astronauts' clean-cut image. Another post tells of the plea of a reverend

to Paine to direct agency technology toward the plights of hunger and poverty. Paine is said to have sagely noted: "It will be a lot harder to solve [these problems] than it is to send men to the Moon."

The series continues through the anniversary of the launch, on 16 July, and will cover the duration of the mission, up until 13 August, when the returning astronauts left quarantine. ■

Visit Nautilus for regular news relevant to *Nature* authors <http://blogs.nature.com/nautilus> **and see Peer-to-Peer for news for peer reviewers and about peer review** <http://blogs.nature.com/peer-to-peer>.

Growing pains

The fledgling European Research Council is struggling against the constraints imposed by the European Commission. It needs to be completely independent.

Two years after its inception, the European Research Council (ERC) seems to be doing well. Set up by its parent body, the European Commission, as the first-ever pan-European agency for funding basic science solely in terms of excellence, it has now run two full rounds of grants and earmarked some 600 projects for funding. But this apparent success masks some troubling issues. The council is facing a number of legal and administrative constraints grounded in the culture of the commission itself.

The issues can be traced back to the ERC's formative years when, after much passionate debate, it was agreed that the council's administration would be handled by an executive agency. This body is legally separate from the European Commission, but is still under its control and bound by its rules and regulations.

This made it possible for the ERC to be born in a speedy and relatively pain-free manner. But it put the agency in danger of being suffocated by the commission's infamous bureaucracy, and by the multitudinous checks and balances that constrain how commission money can be spent. This red tape reflects the culture of mistrust that has haunted the commission since fraud allegations forced commission president Jacques Santer and all his fellow commissioners to resign in 1999, and has often been cited as deterring researchers from taking part in the commission's Framework research programmes.

As was feared, bureaucracy now seems to be stifling the ERC's mission (see page 440). A prime example is that, legally speaking, ERC grant recipients are contractors to the commission — which means they have to fill out onerous time sheets detailing their activities throughout their working day.

The commission should cut through this red tape by embracing the alternative favoured by many academics. It should make the ERC legally independent. Article 171 of the treaty that governs the

European Union specifically allows for the creation of such bodies. An independent ERC would still have to account for the proper use of European Union funds to both the European Parliament and the member states. But it would have the freedom to operate and govern itself in the way it thought best for a body funding basic research.

The key drawback of such autonomy, according to sceptics within the commission and elsewhere, is that it might make the agency more vulnerable to pressure from member states seeking a share of the ERC funds for researchers in their own borders. Such influence would undermine the ERC's mandate of awarding grants solely on the basis of scientific excellence.

These fears are unfounded. Article 171 would allow the ERC to set up any structure it deems appropriate to fulfil its function, and it would be under no obligation to put member-state representatives on its decision-making boards. It is true that member states and the European Parliament would have to vote to agree to such an autonomous structure, and they would also have a vote on the ERC's budget — but this is no different from the ERC's current situation with its executive agency.

When the commission set up the ERC, it deferred until 2010 a decision on whether the council should remain in its current form, or should be restructured once it was established. That decision will be based on an independent review of how the ERC is functioning, which is due to be published this week. This assessment must not be allowed to be buried. The commission now has the chance to make the changes required to ensure that the ERC has a strong future and is a true servant of the scientific community. ■

"Bureaucracy now seems to be stifling the European Research Council's mission."

Beyond the pristine

Earth's disturbed ecosystems have much more to offer than many would give them credit for.

Take a look out of the nearest window. Chances are that the view will show a man-made landscape — a residential neighbourhood, a field of wheat, an overgrazed hillside or a weedy forest full of invasive species. An estimated 77% of Earth's ice-free land has now been substantially altered by human activities, and that proportion is likely to rise.

Scientists and conservationists would do well to pay more attention to the landscape outside the window. At the moment, they tend to concentrate on the remote fraction of Earth that looks more or less

like it did before humans swarmed the globe, and on the protection of those places. And they have good reason to do so: such relatively pristine sites not only harbour much of Earth's biodiversity, they also offer a unique opportunity to learn how nature works in isolation from *Homo sapiens*.

Nonetheless, the majority of Earth that is not pristine provides critical support for human life, in the form of agriculture, development and the 'ecosystem services' that sequester carbon, filter water, build soil, shore up hillsides — and it even creates habitats to shelter threatened species.

About half of that 77% of Earth's surface is in direct use by humans for agriculture and urban development. The other half is marked by past human influences — fragmented forests filled with species from other continents, forest plantations, abandoned grazing lands and so on. Some ecologists have begun to classify portions of the latter half

as 'novel ecosystems' — areas not currently being managed by people, but so changed by human activities that they have become different ecosystems, with different players and a different trajectory from the system they replaced. Many of these places have changed so much that attempting to return them to their historical condition would be prohibitively expensive, if not impossible, especially with climate change added to the mix. Nevertheless, these alternative states are often species-rich, energetically productive and vigorous providers of ecosystem services (see page 450).

Given these realities, more researchers and conservationists need to expand their interests to encompass urban and agricultural ecologies, as well as novel ecosystems in general. Many have already done so. Next month's meeting of the Ecological Society of America in Albuquerque, New Mexico, plans symposia on soil microbial ecology in sustainable agriculture, agroecosystems of the future and urban design. But such work merits increased support from all of the government agencies and non-governmental foundations that fund ecological and conservation work. A good example is the US National Science Foundation's Long Term Ecological Research Network, which includes among its 26 sites two cities — Baltimore, Maryland, and Phoenix, Arizona — where researchers examine how the human and ecological systems in the city interact and change over time.

Likewise, funders that support research into ecosystem services, such as the US Environmental Protection Agency in Washington DC, should put a special emphasis on anthropogenic areas. Some of these places may even warrant protection, as odd as it might seem to create a park around an overgrown orchard or a lake filled with foreign fish.

None of this is to say that anything goes. Exotic species, pollution and mindless development have wreaked widespread havoc in the past, and strong safeguards are still needed to minimize future damage. Moreover, some novel ecosystems — monocultural stands of invasive plants such as leafy spurge, deserts created by nibbling goats and rabbits or bodies of water devoid of all life larger than algae — are universally considered undesirable and are crying out for intervention and restoration. But not all change is bad. Where a reasonably healthy, reasonably diverse ecosystem is providing at least some kind of service, we might be better off to embrace our altered Earth.

Indeed, when society learns to appreciate ecosystems as they are, rather than always yearning to return to an impossible, pristine past, we may be able to make that 77% work for us so well that we never need to disturb the rest. Preservation of the pristine may depend on our understanding and careful use of the worn and grubby. We may even learn to find some charm there. ■

The carbon count

Scientists need better Earth-monitoring tools to see whether climate policies are working.

When the world's nations meet in Copenhagen this December to try to construct a successor to the 1997 Kyoto Protocol on climate change, one major point of discussion will be 'offsets'. These are deals that could help countries meet their targets for reducing emissions by paying for others to absorb greenhouse gases in natural carbon sinks such as rainforests, or by otherwise reducing the threat of global warming.

Any new agreement would presumably build on the existing Kyoto framework that allows certified credits from offsetting projects, such as planting trees, to be traded on the international emissions market. Voluntary carbon offsets are also becoming increasingly popular among businesses and air travellers who want to compensate for the carbon footprint of their activities.

As things stand, unfortunately, the success or failure of any such policy is largely a matter of guesswork: there has never been a global observation network capable of verifying whether the carbon dioxide emissions and offsets reported by individual countries make any sense. Carbon-cycle scientists estimate, for example, that around one-third of the CO₂ from fossil fuels burned globally is taken up by land vegetation. But they have no idea what the precise fraction is, or where the carbon actually goes: *in situ* measurements of biosphere-to-atmosphere carbon fluxes are scarce, and ecosystem inventory data are often unavailable. In addition, monitoring efforts suffered a dire setback on 24 February when NASA's US\$278-million Orbiting Carbon Observatory (OCO) crashed into the ocean minutes after launch.

Any new international climate agreement, whether it emerges at Copenhagen or later, must therefore provide for a much-improved carbon-monitoring infrastructure for verifying its effectiveness. One key element will be satellite observations, which provide large-scale mapping of greenhouse-gas emissions and land-cover changes. NASA should get the support it needs to build a cheaper copy of OCO, which could be launched as early as 2011 (see *Nature* 458, 8; 2009).

But equally crucial will be high-precision, *in situ* measurement of carbon fluxes between soils, vegetation and the atmosphere. The many international agencies that make up the Integrated Global Observing Strategy partnership should implement, without further delay, their 5-year-old plan for an Integrated Global Carbon Observation programme. A good place to start would be to expand FLUX-NET, an existing surface network of some 400 carbon-measurement towers that still has huge gaps, particularly in the tropics.

At the same time, the agencies that comprise the intergovernmental Group on Earth Observations (GEO) could aim to produce globally harmonized data sets on global, national and local scales, using common algorithms, variables and units. GEO, which coordinates efforts to create the Global Earth Observation System of Systems, should also commission scientists to develop an integrated model that stitches all carbon observations together. It should then make these available for use at all levels by scientists and policy-makers alike.

A global carbon-measurement system along these lines should make international climate policies much more solid than they have been in the past. It might reveal that what we are doing is not enough, and that many offset projects fail to deliver. It might expose swindlers and profit-makers in the carbon business. Or it might prove that nature is a stronger ally than we have dared to hope. Whatever the outcome, a serious investment in carbon monitoring will be money well spent. ■

RESEARCH HIGHLIGHTS

Skink or swim?

Science 325, 314–318 (2009)

The sandfish skink (*Scincus scincus*) spends much of its day below the desert surface, shimmying through the sand to escape predators, avoid overheating and sneak up on prey.

Daniel Goldman at the Georgia Institute of Technology in Atlanta and his colleagues used high-speed X-ray imaging to scrutinize the skink's movements as it zipped through sand-grain-sized glass beads. They discovered that the sandfish does not use its limbs to push itself along, but instead tucks them against its body and undulates like a sine wave.

The team then modelled the thrust and drag involved in swimming through the grains. The model correctly predicted that a given frequency of body waves generates the same speed regardless of whether the grains are loose or packed.

For movies, see <http://tinyurl.com/ne9ozh>



C. MATTISON/FLPA

ASTRONOMY

Reionizers spotted

Astrophys. J. 700, 20–48 (2009)

The details of reionization — in which radiation from the Universe's first stars and galaxies ripped apart atomic hydrogen left over from the Big Bang — matter greatly to astronomers. One class of galaxy, comprising those enshrouded in glowing hydrogen gas, is thought to have been important in reionization. But the earliest examples of these are hard to detect because they are so distant.

Using the 10-metre Keck telescopes in Hawaii, Brian Leemaux of the University of California, Davis, and his colleagues found a surprising number of gas-blanketed galaxies in a small patch of sky from a time when the Universe was 1.2 billion to 1.8 billion years old. Even the dimmer galaxies had significant rates of star formation, suggesting that this class plays a crucial part in reionization.

CELL BIOLOGY

Spindle sandwich

Curr. Biol. 19, 1086–1095 (2009)

By squishing cells, researchers hope to better understand the dynamics of their division. Several competing models have been proposed for the mechanics of the cell's spindle — an assembly of microtubules and other structures that stretches across dividing

cells and splits chromosomes between them.

Timothy Mitchison and Sophie Dumont of Harvard Medical School placed dividing cells in an apparatus that gently clamps down on them. The spindles in squashed cells got longer and wider, returning to their normal dimensions after the force was released.

On the basis of their observations, the authors propose that a mechanochemical switch exists at the spindle poles that regulates microtubule length in response to forces within the spindle.

PHYSICS

Scattered showers

Nature Phys. doi:10.1038/nphys1340 (2009)

Scientists had assumed that the scattered size distribution of raindrops hitting the ground resulted from collisions with their neighbours, just like those in clouds. But Emmanuel Villermaux and Benjamin Bossa at the Aix-Marseille University in France determined that the key to this phenomenon is the dynamics of individual drops.

A high-speed camera captured the moment when a single drop fell through the air, flattening like a pancake as it neared terminal velocity. It then inflated into a bag-like shape before splitting into stringy filaments and, finally, bursting into a shower of different-sized droplets (pictured below).

The complete metamorphosis from drop to

droplets occurred more quickly than the typical collision time between drops in clouds.

For a longer story on this research
see <http://tinyurl.com/kop4ww>

CANCER BIOLOGY

Doing more with less

J. Am. Med. Assoc. 302, 276–289 (2009)

Shedding a copy of chromosome 10 helps lethal brain tumours called glioblastomas to thrive by reducing expression of the tumour-suppressor gene ANXA7.

Glioblastomas often have only a single copy of chromosome 10, and extra copies of a region of chromosome 7 that contains the cancer-promoting EGFR gene. Markus Bredel of the Northwestern University Feinberg School of Medicine in Chicago, Illinois, and his colleagues found that expression of ANXA7, which is located on chromosome 10, is also reduced in these cells.

In cell cultures, when ANXA7 protein levels dropped, EGFR levels rose and the tumour-generating potential of these cells increased.

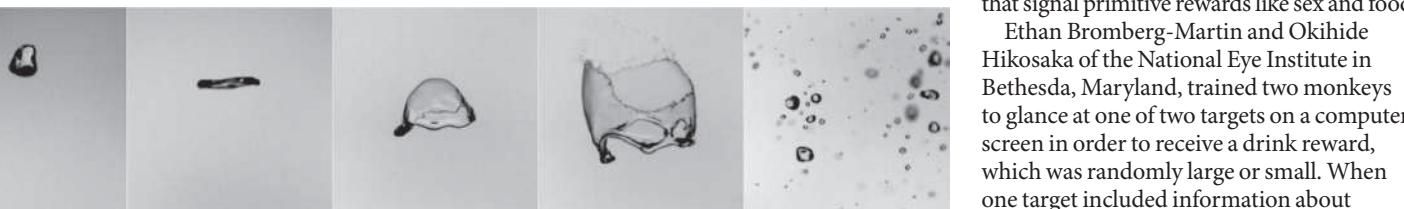
NEUROSCIENCE

Knowledge rewards

Neuron 63, 119–126 (2009)

Monkeys like to know the size of rewards coming their way, and, in the brain, this desire is signalled by the same dopamine neurons that signal primitive rewards like sex and food.

Ethan Bromberg-Martin and Okihide Hikosaka of the National Eye Institute in Bethesda, Maryland, trained two monkeys to glance at one of two targets on a computer screen in order to receive a drink reward, which was randomly large or small. When one target included information about



JOURNAL CLUB

Rex Cocroft
University of Missouri

An ecologist marvels at animals that learn to eavesdrop.

All through college I resisted getting glasses, but I finally succumbed for my first field trip to Peru; I was determined to see everything. Upon arrival, however, I realized that good vision was scarcely enough. One morning, I walked through the forest with an ornithologist, the late Ted Parker, famed for having learned the songs of thousands of birds. Although we saw few of the singers, he knew the source of each fluted phrase, monotonous chant and raucous outburst.

Attending to the signals of other species — visual, auditory and so on — is useful not only to ecologists, but also to the predators that eavesdrop on their prey. In some cases, such behaviour is hard-wired; however, this seems unlikely for species that exploit a wide range of prey.

Recent research has revealed a more flexible strategy. Martinus Huigens of Wageningen University in the Netherlands and his colleagues studied a tiny wasp that parasitizes butterfly eggs (*M. E. Huigens et al. Proc. Natl Acad. Sci. USA* **106**, 820–825; 2009).

The wasp, *Trichogramma evanescens*, learns, after a single experience, to exploit the hosts' chemical-communication system to find and hitchhike on a mated female, disembarking when the butterfly lays her eggs.

Prior research had revealed only one other such case, in the bat *Trachops cirrhosus*, which learns the calls of poisonous and edible frogs (*R. A. Page and M. J. Ryan Curr. Biol.* **16**, 1201–1205; 2006).

These examples suggest that learning which communicative signals to follow may be a common feature of the evolutionary race between predator and prey. It is doubtful that any bat or wasp can retain as many signals as a legendary ornithologist, but it seems that the drive to learn them has a long history.

Discuss this paper at <http://blogs.nature.com/nature/journalclub>

reward size the monkeys preferred to go for that target, rather than be surprised by a randomly sized reward.

The scientists recorded from single neurons in the brain's 'reward' circuitry and found that they fired when the monkeys learned information about the future. This suggests that the act of prediction may be intrinsically rewarding.

EVOLUTION

Safe sex for primroses

Proc. Natl Acad. Sci. USA doi:10.1079/pnas.0904695106 (2009)

Does sex promote pest resistance? Marc Johnson of North Carolina State University in Raleigh and his team tested the susceptibility of 32 species of evening primrose (Onagraceae) to a selection of herbivores. This plant group contains both sexually reproducing species and functionally asexual species.

In the lab, the researchers found that generalist caterpillars, which eat a variety of plants, ate 32% more leaf tissue on asexual than on sexual plants. And in a field experiment, generalist herbivores consumed 64% more leaf tissue on asexual species. However, reduced sexual reproduction seemed to help protect against specialist herbivores.

This, the researchers say, supports the previously untested 'recombination-mating system hypothesis', which holds that sexually reproducing species should have greater resistance to arthropod herbivores.

BIOENERGETICS

Winter wrecked

J. Exp. Biol. **212**, 2483–2490 (2009)

Seabirds wintering off the shores of Newfoundland and Greenland often fall victim to 'winter wrecks', vast die-offs that can cast corpses ashore by the thousand. With the harsh North Atlantic winter, it is difficult to study the mechanisms of these wrecks directly. So Jérôme Fort of the CNRS in Montpellier, France, and his colleagues used a program called Niche Mapper to predict the energy and food requirements of two species from 1 September to 1 March.

They integrated microclimate data with the physiological and behavioural qualities of Brünnich's guillemots (*Uria lomvia*) and little auks (*Alle alle*). The model predicted a sharp increase in energy requirements for both species between November and December that was influenced most by falling air temperatures. Although a preliminary estimate, this energetic bottleneck fits well with the observed timing of winter wrecks in the northwest Atlantic.

PHYSICS

Jet stream

Phys. Rev. Lett. **103**, 024501 (2009)

A millimetre-sized water droplet sitting on a vibrating crystal surface can be forced to form a narrow column more than a centimetre long.

Leslie Yeo and his colleagues at Monash University in Clayton, Australia, created controllable, nozzle-free liquid jets by focusing energy into the water from surface acoustic waves (SAWs): earthquake-like vibrations generated electrically in the underlying crystal.

SAW devices are powerful — a wave 10 nanometres in amplitude can generate surface accelerations 10 million times that of gravity — and inexpensive, already inhabiting most mobile phones. Researchers hope to use them to manipulate fluids at the microscale. They can make liquid drops wobble, slide sideways and break up into a fine mist.



A. DELUCA/CORBIS

PALAEONTOLOGICAL GENETICS

Untraceable Etruscans

Mol. Biol. Evol. doi:10.1093/molbev/msp126 (2009)

Since the time of Herodotus, speculation has surrounded the origins and fate of the Etruscans (responsible for the artwork pictured above), an ancient culture based in central Italy that was strikingly different from its neighbours at the time. Although eventually absorbed by the Roman Empire, the Etruscans' genes survived into the Middle Ages, say Guido Barbujani of the University of Ferrara, Italy, and his colleagues. But between then and now, the line of descent has become much more complicated.

The researchers compared mitochondrial DNA taken from Etruscan remains, from Tuscan bones dating from the tenth to fifteenth centuries, and from modern Tuscans. The same markers were detectable in the two older groups, but not in today's Tuscans, probably owing to migration.

NEWS

Red tape strangles basic research grants

European Research Council finds itself mired in bureaucracy.

EXCLUSIVE

The European Commission's strict rules on finance and administration are hampering efforts by the European Research Council (ERC) to fund scientists, *Nature* has learned. The rigid bureaucracy is causing peer reviewers to desert the grant-review process, and some worry it could threaten the future viability of the council.

The ERC was set up two years ago as a pan-European initiative to fund frontier research judged solely on excellence. At the time, many scientists urged that it should be kept free of the European Commission's notoriously complex bureaucracy (see *Nature* 436, 441–442; 2005). They had hoped that the council would be responsible for its own budget and be allowed to create its own rules on administration, structure and employment. But ultimately it was created as an executive agency run through the commission.

Sources of frustration now emerging include the extensive requirements for documenting the time spent on work funded by the council, plus administrative delays in areas such as expense claims for peer reviewers.

"In general the scientific community is impressed by what the commission has done in establishing the ERC," says Ernst-Ludwig Winnacker, whose term as the council's secretary-general ended on 1 July. "But in time I think they will run out of patience."

On 23 July, a high-level panel led by the former president of Latvia, Vaira Vike-Freiberga, will publish a review of the ERC's structure and governance. It is expected to comment on these problems and make recommendations to address them.

The commission is due to respond officially to the review in October. But speaking exclusively to *Nature*, Janez Potočnik, the European commissioner for research, agrees that some of the current rules are "too rigid" for organizations such as the ERC. "This needs to be looked at clearly, and we are doing it," he says.

If the ERC is to have a strong international reputation, he says, "the commission needs to work more on how to create an autonomous

and accountable institution. And we are doing just that."

"We are of course committed to having the most efficient and appropriate structure for the ERC," adds José Manuel Silva Rodríguez, who heads the commission's research directorate responsible for the ERC.

The vision for the ERC was to give the best European researchers the freedom to explore ideas, through grants similar to those of the US National Science Foundation. So far, the council has selected around 600 projects for funding worth around €900 million (US\$1.3 billion).

The ERC's scientific strategy, priorities and peer-review process are established by its scientific council. Many would like to award researchers a lump sum with as few obligations as possible other than showing the fruits of their endeavours — publications, for instance — at the end of the project, as is done at many other funding agencies. But the ERC must follow the commission's legal rules in awarding contracts. Recipients of awards must negotiate in line-by-line detail how much money they need and on what it will be spent. And they must reach clear milestones with expected outcomes before receiving the next parcel of money.

"You want people to do first-rate frontier research and then you impose contractual obligations on them."

"You want these people to do first-rate frontier research and then you impose contractual obligations on them. The symbolic value of this is obvious," says a member of the ERC's scientific council, who asked not to be named for fear that the commission might baulk at making changes.

Unpaid peer reviewers of grant applications are also hit by red tape. They have to provide proof of their identity, which the commission must authorize, before they can be given any information about the applicant or the proposal they are being asked to review, says Jean-Pierre Bourguignon, chairman of the ERC's peer-review panel for young mathematicians.

"I have never seen anything like this for peer review," says Bourguignon, who heads the Institute of Advanced Scientific Studies (IHES) near Paris. A number of potential peer reviewers "refused to be involved", he adds. "They were very upset."



Janez Potočnik acknowledges the need for change at the European Research Council.

UPPA/PHOTOSHOT

Bourguignon says he has had to write hundreds of e-mails to reviewers apologizing for the rules in an attempt to persuade them to help. He says that out of some 225 reviewers contacted over three grant rounds, two-thirds initially refused to help, and from that group he managed to salvage one-third.

The same problem has been experienced across all the panels, says the science council member. Other researchers who initially signed up to review dropped out in later rounds because the commission took three months to reimburse their travel costs.

Silva Rodríguez says that new procedures are being adopted to streamline processes such as the approval of outside peer reviewers. Travel expenses for overseas experts are now paid in about 20 days, he says.

Teething troubles

"We have wasted an enormous amount of energy trying to solve even the most minor of things that should not be causing any problem. Hopefully this difficult period is now over," says Helga Nowotny, vice-president of the ERC and a professor emeritus in social science at the Swiss Federal Institute of Technology in Zurich. She says the ultimate goal of the scientific council is for the ERC to be administratively and financially independent of the commission.



STRING THEORY GETS NEW LEASE OF LIFE
Baroque field could help explain superconductivity.
www.nature.com/news

M. KULYK/SPL



Winnacker, meanwhile, argues that power needs to shift away from the commission towards scientists. He suggests that the director of the executive agency, who authorizes each funding award, should be a scientist appointed by the scientific council, and that its steering committee should have a majority of scientists. Having scientists in charge, he says, could make it easier to solve some of the problems that the ERC has encountered because there is "room for interpretation" in the commission's rules.

Dieter Imboden, president of EURO-HORCs, a group of the heads of Europe's research councils, agrees. "We wrote a letter to the commission in March about the problems the ERC was having," he says. "Mr Silva Rodríguez met with me to discuss the problems but said that there was nothing he could do. We don't accept this. He does not use the room for manoeuvre he has as the director-general for interpreting the rules."

"All my contacts so far with the scientific community, the European Union member states and research institutions have strongly reassured me that we, the commission and the scientific council, each with their respective responsibility, are on the right track," says Silva Rodríguez.

Potočnik's term ends in October, but he has time to follow up on the results of the ERC review before he leaves. ■

Natasha Gilbert

See Editorial, page 435.

Cuts bite in California

The ten-campus University of California (UC) system — a national star of US public universities — has begun haemorrhaging top researchers as its financial crisis escalates.

On 16 July, the UC board of regents voted to give its president Mark Yudof the power to force university staff to take unpaid leave through a furlough plan. The cuts are meant to help the system offset about a quarter of its US\$813-million drop in state funding this year. The measures would affect 108,000 full-time UC employees and would follow a sliding scale according to salary, with the lowest-paid faculty members drawing 11 unpaid days off — equivalent to a 4% pay cut — and the highest-paid staff drawing 26 unpaid days, or a 10% cut.

"The alternative is massive lay-offs," Yudof told the regents in San Francisco. "We have frankly just run out of money."

Those exempt from salary cuts include student employees, such as graduate students, as well as staff whose salaries are paid wholly from research grants and contracts. Most campuses have dramatically curtailed recruiting, including withdrawing job offers already made to candidates.

Marye Ann Fox, chancellor of the University of California, San Diego (UCSD), said at the regents' meeting that several top researchers had already left the system. In written comments naming some of them, she mentioned Edward Yu, a professor of electrical and computer engineering who is moving to the University of Texas at Austin, and biologist Charles Zuker, said to be moving to Columbia University in New York.

The separate, 23-campus California State University (CSU) system is also struggling — its budget from the state for the new fiscal year is \$584 million less than it was expecting. CSU executives are asking employee unions to agree to a two-day-a-month furlough, which represents a 9.5% pay cut, for 46,000 employees. The California Faculty Association, a key group that represents 23,000 faculty members from the CSU system, is voting on the proposed cut this week, with results likely to be tabulated on 22 July (after *Nature* went to press). If the faculty do not approve, CSU officials say that lay-offs are likely.

At the UC regents' meeting, meanwhile, some scientists argued that young faculty

members will be hit particularly hard by the cuts, and that it is already becoming difficult to recruit and retain them.

Astronomer Mark Krumholz of the University of California, Santa Cruz, says that if the cuts continue for longer than a year, he would consider leaving. "In one year, I have brought in nine times my salary in grants," he says. "If I go elsewhere, that money follows me."

Krumholz spearheaded the writing of a letter sent this month to California governor Arnold Schwarzenegger and members of the state legislature to protest against the cuts. It has been signed by more than 300 UC faculty members who are members of the US National Academies of Science or Engineering or of the Institute of Medicine.

Another signatory at the Santa Cruz campus, astronomer Sandra Faber, warned the regents that continued cuts would lead to serious harm. "The university is the most powerful economic engine in the state," she said. "Disinvesting in the

University of California at this time is like eating our seed corn."

But the state continues to face a \$26-billion budget deficit; it must balance its books for each fiscal year, which began on 1 July. "I don't really see a light at the end of the tunnel," Yudof said last week. With Russell Gould, chairman of the UC board of regents, he has announced plans to form a Commission on the Future of UC to examine the university's continued existence, its services and its funding model.

The cuts have even triggered fighting within the UC system itself, with some at the biggest campuses suggesting turning on the smaller ones. In a 15 June letter that made headlines last week, the chairman of UCSD's sociology department, Andrew Scull, and 22 other department chairs called on the regents to make the UC campuses at San Diego, Berkeley, Los Angeles and San Francisco into elite research institutions; other campuses would be downgraded to teaching institutions, with presumably smaller budgets.

As *Nature* went to press, the state legislature and the governor had agreed a plan to close the budget gap, involving some \$15 billion in proposed cuts. ■

Erika Check Hayden and Rex Dalton

For a video of UC protests over the cuts, see <http://tinyurl.com/ucprotest>

Insuring against climate

Negotiators push for policies to help weather natural disasters.

Farmers in the Ethiopian village of Adi Ha have been busy sowing fresh crops of grain in recent weeks, as is customary when their maize crops struggle because of drought. But this year, they have a second backstop against hunger: insurance.

In Adi Ha, farmers can pay a one-time fee of US\$5 to \$30 to cover their crops of the grain teff, used to produce the flatbread injera. Depending on how much rain falls on this particular swath of the northern highlands in August and September, policies pay out up to five times the premium. The arbiter will be a satellite, marking the first time that scientists have used space-based observations to fashion contracts at the level of individual farmers.

Unlike standard crop insurance, which requires on-the-ground audits, any payments will be distributed automatically according to a set formula, helping villagers to keep food on the table and buy seeds to start over again next year. "Teff is insurance for these farmers, so by insuring teff we are strengthening their insurance," says Daniel Osgood, a researcher who helped develop the policy at Columbia University's International Research Institute for Climate and Society in New York. Oxfam America and the insurance giant Swiss Re are also involved.

Adi Ha is one of dozens of pilot 'index insurance' programmes that are popping up



throughout the developing world as governments, non-profit groups and aid agencies look for ways to help poor communities — and in some cases countries — cope with natural disasters. The idea has now taken root in the United Nations' climate talks; many delegates will be pushing to incorporate insurance policies into a deal at the climate summit in Copenhagen this December.

Leading the effort is the Munich Climate Insurance Initiative (MCII), whose members include reinsurance companies, research groups and agencies such as the United Nations Development Programme. The MCII's proposal combines incentives to help communities prepare for natural disasters in a two-tiered programme: a climate-insurance pool would pay for major disasters in developing countries, and a second tier would promote public-private insurance systems to pay for broader implementation of index and other types of insurance programme.

All told, the proposal could cost around \$10 billion annually, says Koko Warner of the United Nations University in Bonn, Germany, who manages the MCII initiative. "The hook would be that in order to qualify for insurance programmes, countries would have to show that they are actively engaged in risk reduction," she says. "Developing countries agree to reduce their risk, and then developed countries would



provide an insurance package."

Warner says that references to climate insurance go all the way back to the original United Nations Framework Convention on Climate Change signed in Rio de Janeiro in 1992, but only in the past couple of years has the concept really been integrated into international climate negotiations. In addition to fairly widespread support among developing nations, the European Union has said that it is willing to explore the idea, Warner says, and the steadfast opposition from the United States during the George W. Bush administration has softened under President Barack Obama.

Perhaps the biggest stumbling block is the hard-line position being taken by island

US Congress revives hydrogen vehicle research

US funding for hydrogen-fuelled transportation research got a boost on 17 July as the House of Representatives voted to restore \$85 million to the research budget. The administration of President Barack Obama had proposed cutting the funds altogether.

In May, energy secretary Steven Chu sparked an uproar when he proposed slashing current spending on research into hydrogen-based energy technology by 60%, from \$168 million this fiscal year to \$68 million in 2010, and cutting funding entirely

for work on hydrogen vehicles. Former president George W. Bush made hydrogen transportation a cornerstone of his energy research strategy, but Chu said biofuels and batteries offer a better short-term pathway to reducing oil use and greenhouse-gas emissions.

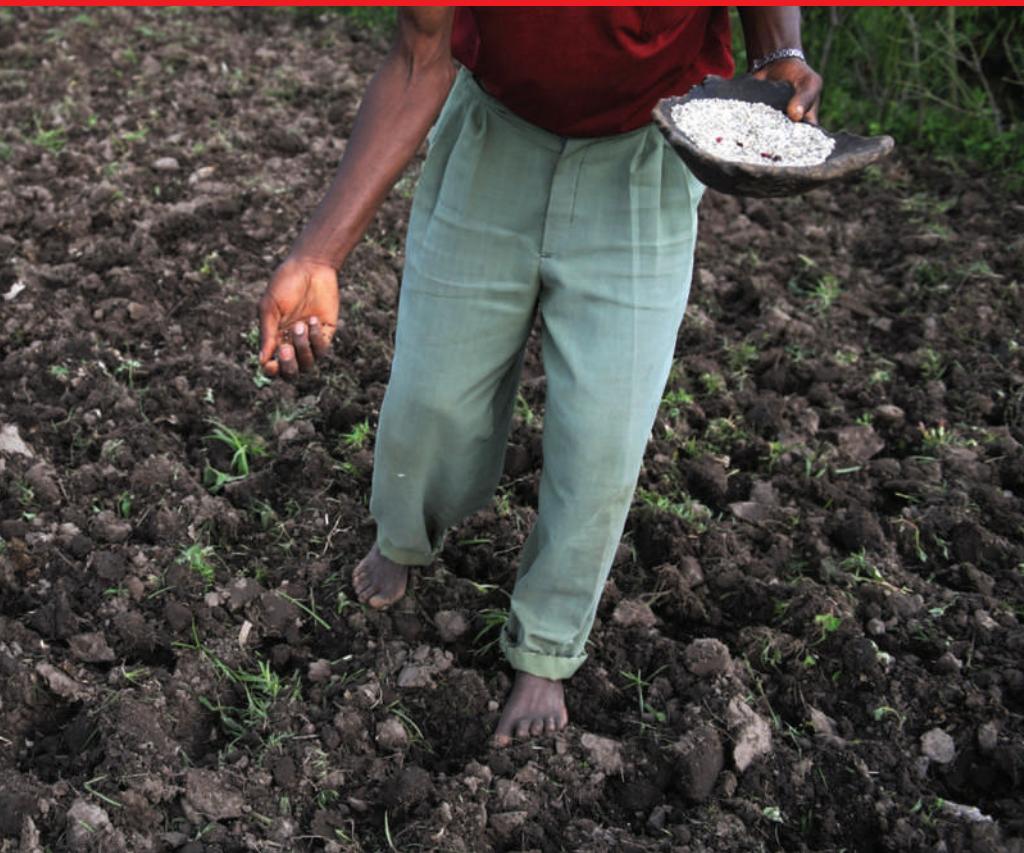
Advocates both among scientists and on Capitol Hill have rushed to defend the hydrogen programme in recent weeks. It seems to have worked: the House included a total of \$153 million for hydrogen-energy research in its version of the 2010 energy and water spending bill.

In the Senate, appropriators have provided \$190 million for hydrogen research — a 13% increase over the base budget for 2009 — although the full Senate has yet to take up the legislation. A final bill is unlikely to come for another few months, but some level of funding for hydrogen vehicle research is likely to survive.

Also last week, a National Research Council (NRC) panel weighed in on the debate with a preliminary report on the FreedomCAR and Fuel Partnership, a research consortium involving industry and government. The

NRC committee endorsed the general thrust of the transportation research agenda of the Department of Energy (DOE) but said it is concerned about efforts to scale back work on hydrogen-fuelled transport. Citing the long-term potential of hydrogen fuel cells, the panel said it is not yet clear which vehicle technologies will prevail in the market.

"There was no disagreement on the DOE's approach to put more emphasis on nearer-term technologies, but we felt that the long-term, high-risk, high-payoff



CLIMATE TARGETS FOR SHIPS DEFERRED

Pollution to be cut back but no deal on carbon dioxide.
www.nature.com/news

B. KILIC/AFP/GETTY

Catastrophe Risk Insurance Initiative. Launched in 2007 with \$47 million in funding from several international donors, the index insurance pool now provides hurricane and earthquake insurance to 16 Caribbean nations.

Much as the Ethiopian policy is tied to rainfall, the Caribbean policies are based on observations of wind speed and earthquake intensity. That saves money on site audits and speeds up the process, providing money immediately after a crisis when it is needed most. "It gives them a bit of breathing room," says Simon Young, the Washington DC-based head of the non-profit firm Caribbean Risk Managers, which manages the programme.

So far, the Caribbean programme has paid out nearly \$1 million for an earthquake that affected St Lucia and Dominica in 2007, and \$6.3 million to the Turks and Caicos Islands after Hurricane Ike last year. Young says that their models try to take into account factors such as building codes and other preventative action, which should lower premiums as well as lessen damage during a storm.

Insurance advocates acknowledge that spreading insurance tools around the globe would benefit developing nations regardless of global warming, as illustrated by the Caribbean initiative. But fears about increased droughts, floods and more severe weather that could be associated with global warming have added momentum.

"Climate change is more or less a new impulse for promoting this," says Thomas Loster, chairman of the Munich Re Foundation, a non-profit philanthropic branch of the German reinsurance giant. "But of course we should have done it 20 years ago."

Jeff Tollefson

Farmers are looking for backup plans in case their crops fail.

R. SCHMIDT/AFP/GETTY

nations. The Alliance of Small Island States has proposed a mechanism that is similar to the MCII proposal but with one key difference: its members want outright compensation, rather than just insurance, for long-term problems associated with issues such as ocean acidification and rising sea levels.

Responsible action

The word 'compensation' raises concerns in industrialized nations, who don't want to sign a blank cheque, but the alliance isn't backing down. Many see the language as a warning to industrial nations regarding the costs of

inaction. "If you are one of these low-lying atolls in the Pacific, would you say 'thank you very much' to a deal that submerges your island over time?" asks M. J. Mace, a negotiator for the Federated States of Micronesia. "If there's a deal, it's got to address impacts, one way or another."

One potential compromise under discussion would be to include an insurance mechanism in whatever deal is struck and then acknowledge the long-term compensation issue in a more symbolic manner.

To date, the best model for large-scale multilateral insurance may be the Caribbean

activities should not be abandoned, in particular those related to hydrogen fuel cells," says Vernon Roan, a retired professor from the University of Florida in Gainesville who chaired the panel.

Pat Davis, who manages the DOE's Vehicle Technologies Program in Washington DC, says the department requested the report to update its vehicle research plans. He called the report "highly favourable" in general, but acknowledged that the administration has a different view of hydrogen research.

Hydrogen fuel cells combine hydrogen and oxygen to generate electricity, producing only water

vapour in the process, and they have already powered prototype vehicles. Fuel cells are expensive, however, as would be the infrastructure required to support large numbers of hydrogen-powered vehicles on the roads. And although renewable energy sources could be used to produce hydrogen, at present it is generally made from natural gas in a process that also produces greenhouse gases.

Nonetheless, hydrogen's advocates say they are making progress on all these fronts, in part thanks to support from within the DOE itself. Byron McCormick, who headed the fuel-cell programme at General Motors until he retired in

January, was a member of the DOE's own technical advisory committee on hydrogen fuel cells. He resigned this spring, however, frustrated because Chu had not reached out to the committee before proposing to slash hydrogen research funding.

"I decided that I had better things to do with my time than communicate with somebody who didn't seem too interested," McCormick says. He points to ongoing research programmes in Europe and Japan and says he found it particularly "disconcerting" that the Obama administration would make such an assessment, despite its emphasis on clean energy. "It strikes me as rather bizarre that

the United States would be the only country backing away from such initiatives," he says.

Patrick Serfass, a spokesman for the National Hydrogen Association in Washington DC, says the DOE's proposal to slash hydrogen research surprised both businesspeople and researchers. Hundreds of pilot fuel-cell vehicles are already on the roads, and major car-makers are preparing to roll out hydrogen fuel-cell vehicles by the middle of the next decade, he says.

"This decision was not really made with a lot of outside opinion or outside input from the industry," Serfass says.

Jeff Tollefson



**TSUNAMI FORECAST
IN REAL TIME**
New Zealand quake yields
unique opportunity.
www.nature.com/news

K. HOKUSAI

Psychiatry manual revisions spark row

US psychiatrists divided by claims of secrecy and scientific overreach.

The question of how best to revise the 'bible' of American psychiatry once again has tempers flaring. The manual, the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM*), is significant because it is used to determine clinical diagnoses, insurance reimbursements and research agendas throughout the United States, and is often used as a reference in other countries.

Planning for the upcoming fifth edition of the manual (*DSM-V*) began in 1999, but as work has picked up during the past year, critics have alleged that the process has been too secretive, and that working groups have been pushed to meet an unrealistic 2012 publication date. Some, including the architects behind the last two editions of the *DSM*, also complain that project leaders are pushing for the premature inclusion of changes meant to incorporate recent genetic and neurobiological advances, before they are ready for the clinic.

Supporters, meanwhile, say that the changes will not be too drastic, and are meant to make the manual more flexible for future revisions.

Even light tweaking of definitions in the *DSM* can bring about radical changes in psychiatric practice, as Allen Frances, professor emeritus of psychiatry at Duke University in Durham, North Carolina, wrote in a recent commentary in *Psychiatric Times*. Frances, chairman of the committee that produced the fourth edition of the manual in 1994, acknowledged that changes in the definitions of autistic disorder and attention-deficit/hyperactivity disorder made then may have contributed to the recent surge in diagnoses of these conditions.

In addition, the manual sometimes has an outsized influence on research directions, says Steven Hyman, provost of Harvard University and a member of the *DSM-V* task force that is overseeing the revision. Hyman became interested in reforming the *DSM* when he was director of the National Institute of Mental Health in Bethesda, Maryland, and witnessed the control that it exerted over grant review panels. "I was spending taxpayer money on grants that were being forced into categories that might or might not conform to nature," he says.

The latest revisions come as financial ties



For the fifth edition, David Kupfer hopes to expand the remit of the psychiatry manual.

between prominent psychiatrists and pharmaceutical companies are being closely scrutinized. A 2006 analysis of potential conflicts of interest among those who participated in the last revision showed that 56% of panel members had financial links to the pharmaceutical industry (L. Cosgrove *et al.* *Psychother. Psychosom.* 75, 154–160; 2006). For the *DSM-V*, the American Psychiatric Association, which publishes the manual, vetted potential members of working groups under a new conflict-of-interest policy. But the process introduced delays, and working groups were not finalized until 2008.

Then word broke last July that working group members had signed non-disclosure agreements, agreeing to refrain from distributing pre-publication materials or divulging the content of group discussions pertaining to the rewrite. The agreements were intended to prevent members from publishing material to be used in the *DSM-V*, says Darrel Regier, vice-chair of the *DSM-V* task force

"There just hasn't been time to do this in an organized way."

and director of the American Psychiatric Association's research division. But Robert Spitzer, a professor of psychiatry at Columbia University in New York who oversaw the *DSM-III*, says that his request to look at minutes of a *DSM-V* meeting was denied because of these confidentiality agreements.

Since then, *DSM-V* working groups have begun posting regular summaries of their activities online. "All of us have been encouraged to be as public as we can be," says William Carpenter of the Baltimore-based University of Maryland School of Medicine and chair of the working group on psychotic disorders. "But just to have [the confidentiality agreements] — that never would have been considered in the previous revisions," says Spitzer.

From the start, Hyman and *DSM-V* chairman David Kupfer have also planned to change how the *DSM-V* evaluates mental disorders. Rather than relying strictly on categorical diagnoses — one either has depression or does not, for example — they have pushed to add 'dimensional' criteria to ascertain to what extent a person is depressed. Such criteria could also address similarities among different disorders, reflecting, for example, neuroimaging studies that suggest multiple anxiety disorders can affect the same region of the brain.

For this reason, Carpenter and others have said the *DSM-V* will represent a "paradigm shift" — an expression that alarmed critics, who say the science behind such dimensional assessments is not yet ready to be incorporated into clinical assessments. In March, Duke University psychologist and epidemiologist Jane Costello resigned from the working group on child and adolescent disorders after receiving a memo from Kupfer and Regier about including the dimensional approach. Adding these assessments would require a great deal of extra research, she says, at a time when working groups were already behind schedule for their 2012 publication deadline. "There just hasn't been time to do this in an organized way," she says. "This is a huge job."

Hyman says that the changes will not necessarily be so drastic, and could take the form of a few additional, and optional, diagnostic criteria without replacing the old methods. ■

Heidi Ledford

Regulators face tough flu-jab choices

Rich countries' pandemic strategies may cause vaccine shortages elsewhere.

Imminent decisions on a strategy for H1N1 pandemic flu vaccination in the United States could leave other countries short of vital doses if it elects not to follow World Health Organization (WHO) advice on vaccine formulation.

The United States is the biggest buyer among a group of rich countries whose combined orders for vaccine against the H1N1 2009 virus could potentially tie up most of the world's pandemic vaccine production capacity for 6 months or longer, so depriving other countries of vaccine.

To counter this prospect, the WHO recommended on 13 July

that countries use shots that contain adjuvants, chemicals that boost the immune system's response to a vaccine. This allows smaller amounts of antigen — the molecule that stimulates the immune response — to be used in each dose, boosting the overall amount of vaccine available from existing production capacity and allowing orders to be filled more quickly.

The United States' global responsibility to consider dose-sparing strategies is briefly alluded to in the minutes of a mid-June US National Biodefense Science Board meeting, released on 17 July: "Federal decision-making will affect not only the 300 million Americans who depend on the government to support the public health system but also people all around the world."

The United States has certainly kept open the option of using adjuvants. It has already allocated almost US\$2 billion for antigen and adjuvant to provide every American with up to two doses of vaccine. That sum includes orders of \$483 million for Novartis's MF59 adjuvant, and \$215 million for GlaxoSmithKline's AS03 adjuvant.

But although Canada and many European countries are set to use adjuvanted pandemic flu vaccines, the United States may do so only as a last resort. "All things being equal, an unadjuvanted vaccine is often just fine in terms of giving protection against influenza virus," Anne Schuchat, director of the National Center for Immunization and Respiratory Diseases



A fistful of vaccines: but will there be enough to go round?

at the Centers for Disease Control and Prevention in Atlanta, Georgia, told a media briefing on 17 July.

"Adjuvant use would be contingent upon showing that it was needed or clearly beneficial," added Jesse Goodman, acting chief scientist and deputy commissioner of the Food and Drug Administration (FDA). "But we want them on the table in case there are issues where they might be needed to protect people in this country."

If there is significant genetic drift in the virus, for example, adjuvanted vaccines are better able to handle such strain variations. And early attempts at

pandemic vaccine manufacture are so far producing two to four times less antigen than seasonal flu strains, raising the threat that the world's production capacity is actually much less than was hoped.

If each shot of pandemic flu vaccine contains 15 micrograms of antigen — the dose used in seasonal flu — and no adjuvant, annual global capacity stands at about 876 million doses, according to the WHO. But as virtually no one is immune to the virus, most experts say that each person will need two doses, immediately halving that capacity. Moreover, higher doses of antigen may be needed to get an adequate response, further reducing capacity. Using adjuvants would boost annual capacity — to more than two billion doses in some WHO projections.

Europe is well placed to quickly authorize adjuvanted pandemic vaccines. Since 2003,

the European Medicines Agency (EMEA) has had a fast-track approval system in which manufacturers can prepare 'mock-up dossiers' — vaccine registration applications that use non-pandemic viral strains but for which pandemic strains can subsequently be substituted. GlaxoSmithKline and Novartis already have mock-up dossiers in place for the H5N1 avian flu virus, and plan to file H1N1 substitutions by the end of July.

Although the EMEA requires the companies to provide new clinical testing and data as they roll out their products, the product itself can be approved in five

days if the agency is satisfied that the extrapolation to the new strain is valid, says Martin Harvey-Allchurch, a spokesman for the EMEA. In contrast, the United States has never licensed an adjuvanted flu vaccine, and has no fast-track system in place, although the FDA can give emergency authorization for new vaccines. The regulators are also mindful of political and public concerns about mass vaccination of the population, given that a vaccination programme in 1976 against a new strain of swine flu caused neurological side effects in about 1 in 100,000 people, and killed 25. Modern flu vaccines, however, have a very good safety record.

The WHO's Global Advisory Committee on Vaccine Safety says "no significant safety concern or barriers" exist to using adjuvanted pandemic H1N1 vaccines. But regulatory agencies may have to approve pandemic vaccines — both adjuvanted and non-adjuvanted — without all the data they would normally require, warns Marie-Paule Kieny, the WHO's vaccine research director. Some preliminary clinical and safety data may be available by September, when flu cases could surge in the Northern Hemisphere, but complete data for adults are unlikely to be available until the end of December, and not until February 2010 for children. Regulators would accompany pandemic vaccine rollouts with parallel clinical trials, and, as in any mass-vaccination campaign, extensive surveillance would monitor for any adverse side effects.

Declan Butler

See also www.nature.com/swineflu

Scientists strive to boost US-Cuban collaboration

A drive to increase scientific exchange between the United States and Cuba is off to a slow start. In the past four months, Cuban officials have cancelled two planned trips of top US scientific leaders to the island nation.

Citing other visitors and events that took up their time, the officials have turned down requests for scientists to enter the country organized by the American Association for the Advancement of Science (AAAS) and the New America Foundation non-governmental organization, both based in Washington DC. In April, the administration of US President Barack Obama said it would work to improve relationships between the two countries, including promoting the “freer flow of information”.

The organizers, who have had the trips in the works since before Obama took office, remain hopeful that a delegation might visit Cuba this autumn, says Lawrence Wilkerson, who was chief of staff to former Secretary of State Colin Powell and is working on a New America initiative aimed at Cuba. The delegation is expected to address topics such as tapping Cuba's strengths in biotechnology, pharmaceuticals and studies involving hurricane research, food production and salt-resistant crops.

“Of course we would like more scientific exchange,” says Miguel Abad Salazar, a researcher at the BIOECO conservation facility near Santiago in eastern Cuba.

Travel restrictions remain a major stumbling block for US–Cuban collaboration. For instance, US scientists seeking to travel to Cuba can't use federal funds without special government permission. And any US scientist travelling to Cuba must get a licence from the treasury department to spend US dollars there, even if funds come from the private foundations that typically pay for such trips.

It has also been nearly impossible for Cuban scientists to come to the United States; one immediate barrier is the US\$150 non-refundable fee for a visa application.

During the Obama administration, however, a handful of Cuban scientists have visited the United States, and US scientists have been increasingly venturing to Cuba. Observers say that the exchanges reflect a

growing thaw in bilateral relations, which began before Obama's election.

In May, for instance, David Winkler of Cornell University in Ithaca, New York, went to Cuba to teach an ornithology course to about two dozen scientists at a BIOECO meeting. He went as part of Cornell's Neotropical Conservation Initiative, coordinated by Eduardo Iñigo-Elias, who has been studying in Cuba for years. “The students and scientists were as well trained as anywhere in Latin America,” says Winkler. “They would be great ambassadors to work on research projects in other countries.” His group hopes to develop such an exchange programme.

These individual exchanges, rather than a coordinated governmental programme, should be the wave of the future, says Peter Feinsinger, a wildlife conservationist who has visited Cuba about a dozen times in the past

six years to train biologists. “I favour a scientific grass-roots initiative,” says Feinsinger, who works for Northern

Arizona University in Flagstaff but is largely funded through the Wildlife Conservation Society in New York. “I think this will happen naturally.”

US researchers often partner with colleagues in other countries to do fieldwork in Cuba. For instance, Kam-biu Liu, a geographer at Louisiana State University in Baton Rouge, collaborates with Matthew Peros, an ecologist at the University of Ottawa in Canada, to acquire sediment samples from Cuba to track hurricane history in the region. Peros has to perform the isotopic analysis on the cores for Liu, who cannot use US funds for the research.

None of the US-based funds that flow into the Inter-American Institute for Global Change Research — the organization based outside São Paulo, Brazil, that funds the hurricane work — can similarly be used for work in Cuba. “I have to invent constructs to fund these projects” with funds from other sources, says Holm Tiessen, the institute's director.

Observers hope that more aggressive efforts to ease US–Cuban relations will be forthcoming as more people fill key jobs in the US Department of State.

Rex Dalton

Cash boost for mapping the human brain

The US National Institutes of Health last week launched an initiative to map the wiring of the human brain. The Human Connectome Project will provide \$30 million over five years for work detailing the connections between major brain subdivisions in healthy adults.

Neuroscientists increasingly view the human connectome as crucial to understanding mental function and disease (see *Nature* 457, 524–527; 2009). The project demonstrates commitment to this nascent research field, and will generate “a new class of data for human neuroscientists”, says its leader Michael Huerta of the National Institute of Mental Health.

Researchers will try to link brain connectivity to genetics and behaviour, by collecting DNA samples, demographic information and behavioural data from their subjects. The initiative also aims to improve non-invasive brain-imaging technologies.

ExxonMobil invests in algae for biofuel

Oil and gas company ExxonMobil, whose chief executive Rex Tillerson called the idea of ethanol as a biofuel “moonshine” in 2007, last week announced a US\$600-million research alliance to develop biofuels from photosynthetic algae.

The multi-year project sees Exxon team up with Synthetic Genomics, the biotechnology company in La Jolla, California, co-founded by Craig Venter, which numbers another oil and gas giant, BP, among its investors.

Synthetic Genomics will receive \$300 million — more, if deemed successful — to develop high-yielding algal strains and their large-scale cultivation. Exxon is spending an equal sum internally on engineering and manufacturing expertise to support the research.



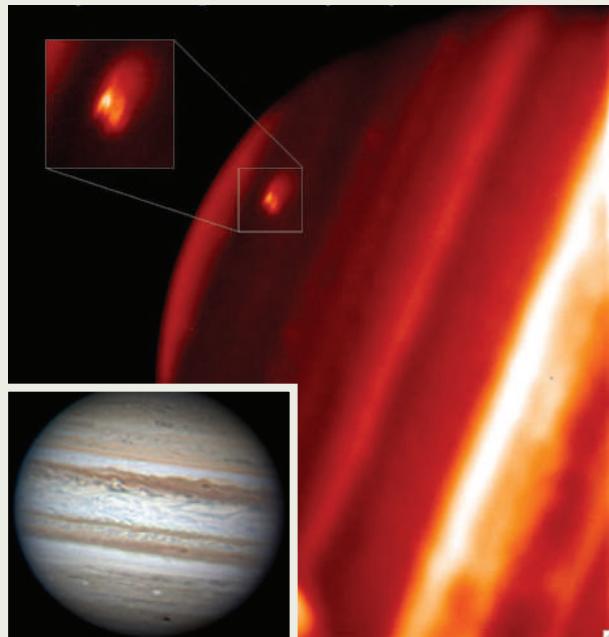
Green power: these tubes of algae are geared up to produce biofuel.

Jupiter takes a hit

An amateur astronomer has spotted the impact of an unknown object on Jupiter. The dark spot near Jupiter’s southern pole (pictured, inset) was detected by Anthony Wesley of Murrumbateman, near Canberra, Australia, on 19 July. Shortly afterwards, word of the finding spread quickly via e-mails and the Internet.

Follow-up observations by the Keck infrared telescope (pictured) and by NASA’s InfraRed Telescope Facility, both on Mauna Kea in Hawaii, rule out a storm, according to Glenn Orton, a planetary scientist at the Jet Propulsion Laboratory in Pasadena, California. “This looks like nothing on Jupiter that is occurring naturally,” Orton says.

The impact took place almost exactly 15 years after the first fragments of the comet Shoemaker-Levy 9 slammed into the giant planet.



A. WESLEY; P. KALAS, UCB; M. FITZGERALD, LLNL/UCB; F. MARSH, SETI INST./UCB; J. GRAHAM, UCB

Italian court sidesteps stem-cell challenge

An Italian court has informed scientists that they have no right to demand a halt to a call for research proposals that specifically excludes the use of human embryonic stem cells — even though their use is legal.

The researchers had objected when the exclusion was added by politicians to a text agreed by a committee of scientific experts (see *Nature* 460, 19; 2009). They argued that it contravenes a constitutional right to freedom of research.

The court said that only research institutes, not individuals, were eligible to object. The scientists are appealing against the decision in a higher court.

Copernicus honoured in periodic-table addition

The name of element 112 should be copernicium (Cp), its discoverers proposed last week. Sigurd Hofmann’s team at the GSI Helmholtz Centre for Heavy Ion Research in Darmstadt, Germany — who first created atoms of the element in 1996 — picked the name to honour astronomer Nicolaus Copernicus, famed for his heliocentric theory of the Solar System.

The superheavy element, previously known by the placeholder ununbium, was

officially recognized by the International Union of Pure and Applied Chemistry (IUPAC) in May this year.

IUPAC confirmation of the new moniker awaits a six-month discussion period to iron out potential confusions. For instance, chemists already use Cp as shorthand for the cyclopentadienyl fragment C_5H_5 .

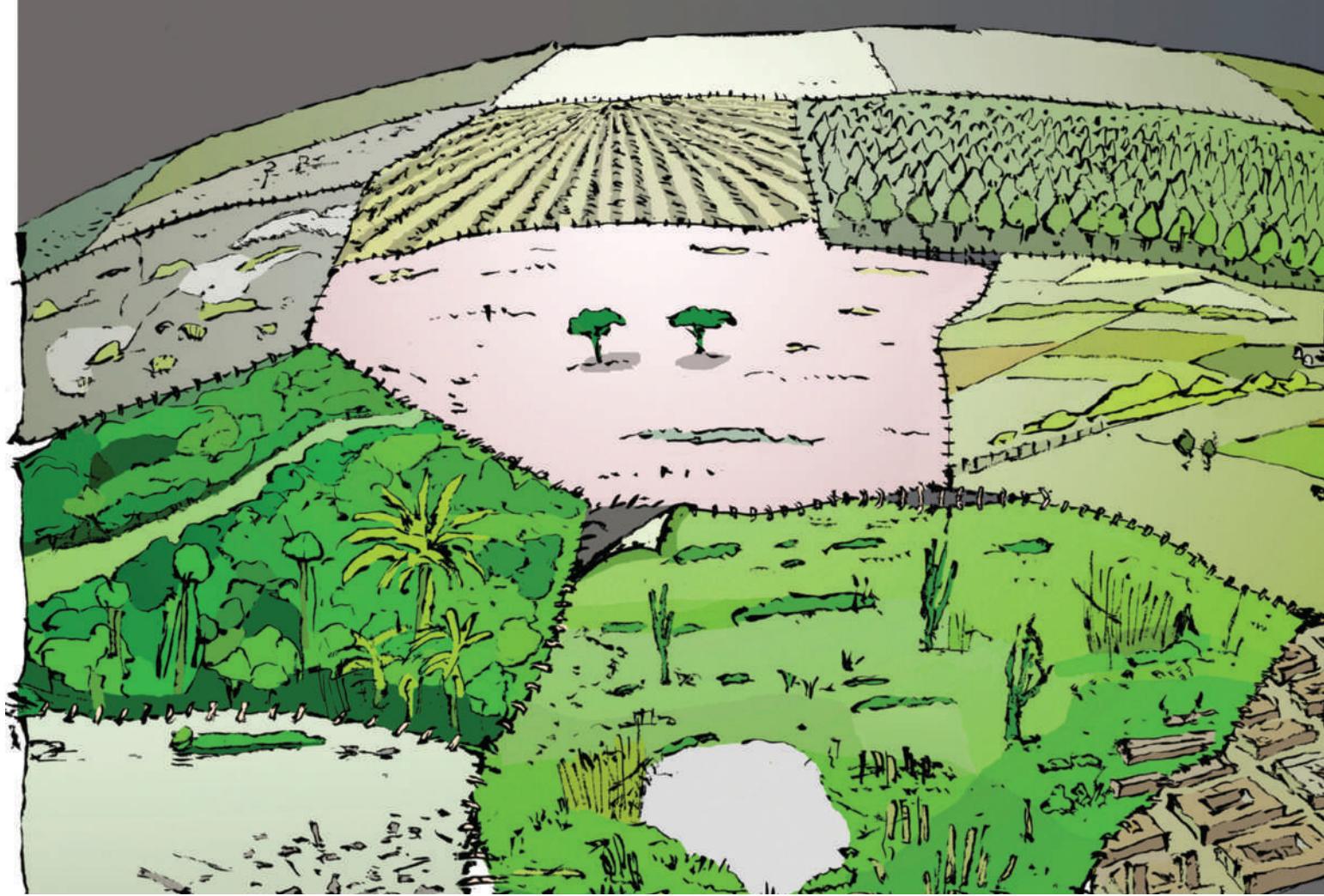
NOMINATE JAPAN'S BEST MENTORS

The *Nature* Awards for Mentoring in Science are annual prizes that have been awarded by *Nature* since 2005 in recognition of excellence in the nurturing of young scientists.

The 2009 awards will be held in Japan, the first time that they have honoured mentors from an Asian nation. Nominations are invited for outstanding mentors from any scientific discipline based in Japan, in two categories: mid-career (up to 59 years of age) and lifetime achievement (60 years of age and over).

Nominations can come from current or former students or colleagues of the nominee from anywhere in the world, and must be supported by two additional people mentored at different times during the nominee’s career. The awards, worth ¥1.5 million (US\$16,000) each, will be presented in December 2009 at a ceremony at the UK ambassador’s residence in Tokyo.

Nominations opened on 22 July, and will close on 25 September 2009. Applications may be made in either Japanese or English. Further details and nomination forms are available for download from www.natureasia.com/en/mentor (English) or www.natureasia.com/japan/mentor (Japanese).



RAGAMUFFIN EARTH

A small group of ecologists is looking beyond the pristine to study the scrubby, feral and untended. **Emma Marris** learns to appreciate 'novel ecosystems'.

Joe Mascaro, a PhD student in a T-shirt and floral print shorts, is soaking in the diversity of the Hawaiian jungle. Above, a green canopy blocks out most of the sky. Aerial roots wend their way down past tropical trunks, tree ferns and moss-covered prop roots to an understorey of ferns and seedlings. The jungle is lush, humid and thick with mosquitoes. It is also as cosmopolitan as London's Heathrow airport.

This forest on Big Island features mango trees from India (*Mangifera indica*); *Cecropia obtusifolia*, a tree with huge star-shaped leaves from Mexico, Central America and Colombia; rose apples (*Syzygium jambos*) from southeast Asia; tasty strawberry guava (*Psidium cattleianum*) from the threatened Atlantic coast of Brazil; and a smattering of Queensland maples (*Flindersia brayleyana*) from Australia. It also has candlenuts (*Aleurites moluccana*), a species that humans have moved around so much that its origins have become obscure. There is at least some native Hawaiian representation in the form of hala, or screwpine (*Pandanus*

tectorius), which is pictured on the crest of Punahoa School, where US President Barack Obama studied. There are no Hawaiian birds here though. Mascaro sees plenty of feral pigs, descendants of those brought by settlers from other parts of Polynesia or from farther afield. The soil is black and rich. Mascaro likes it here.

Most ecologists and conservationists would describe this forest in scientific jargon as 'degraded', 'heavily invaded' or perhaps 'anthropogenic'. Less formally, they might term it a 'trash ecosystem'. After all, what is it but a bunch of weeds, dominated by aggressive invaders, and almost all introduced by humans? It might as well be a city dump.

A few ecologists, however, are taking a second look at such places, trying to see them without the common assumption that pristine ecosystems are 'good' and anything else is 'bad'. The non-judgemental term is 'novel ecosystem'. A novel ecosystem is one that has been heavily influenced by humans but is not under human management. A working tree

plantation doesn't qualify; one abandoned decades ago would. A forest dominated by non-native species counts, like Mascaro's mango forest, even if humans never cut it down, burned it or even visited it.

No one is sure how much of Earth is covered by novel ecosystems. To help with this article, *Nature* asked Erle Ellis at the University of Maryland, Baltimore County, who produces maps of ways that humans use Earth, to take a stab at quantifying it. Defining novel ecosystems as "lands without agricultural or urban use embedded within agricultural and urban regions", Ellis estimates that at least 35% of the globe is covered with them (see map, overleaf). Their share of the planet will probably expand, and many ecologists think that these novel ecosystems are worthy of study and, in some cases, protection.

For one thing, some novel ecosystems seem to provide a habitat for native species — sometimes crucial habitat, if all that the species originally had is gone. They also often do a good



ILLUSTRATIONS BY JANUSZ KAPUSTA

job of providing 'ecosystem services', those things that nature does that benefit humanity, such as filtering water in wetlands, controlling erosion on hillsides, sequestering carbon from the atmosphere and building soil. Provision of ecosystem services is a popular argument for preserving intact ecosystems, but many of its advocates blanch a little when it comes to making the same case for these 'weedy' areas.

Mascaro actually prefers novel ecosystems to some native ones that are so vulnerable to damage by humans that they require intense management to maintain in their 'pristine' state. He sees the latter as museum-piece parks. "Do we value the fact that nature contains a list of things that were there 1,000 years ago, or do we value it because it has its own processes that are not under human control?" Mascaro asks. For him, the value is in the processes.

Watching such processes unfold has scientific merit to many researchers. Novel ecosystems are often ideal natural experiments for studying things such as community assembly — how species find their way to a place and which species become permanent residents — and evolution of species in response to one another. In essence, it takes a dynamic ecosystem to study ecosystem dynamics, and these novel ecosystems are the planet's fastest movers. Mascaro bets that all the rules of thumb and general relationships developed over the years

by ecologists working in 'intact' or 'historical' ecosystems will probably also apply in these new assemblages, but no one knows for sure, because no one has studied them much.

There are some questions about the ways in which things might be different in novel ecosystems. Will landscape types remain the same, with forests replacing forests and grasslands replacing grasslands? Will novel ecosystems evolve faster? Will they be dominated by one species, as many who study invasive species fear? Will species composition oscillate wildly for decades or even longer? "We can't know except to observe it," says Mascaro.

Havens of biodiversity?

One of the first researchers to see the importance of the scrubby parts of Earth was Ariel Lugo, a forest-service ecologist in Puerto Rico. In 1979, Lugo was managing researchers who were measuring the ground covered by trees within pine plantations that were not being actively managed. His technicians came back to headquarters sweaty and discouraged. "They said that they couldn't measure the trees without clearing all the new undergrowth," says Lugo. "They said it was impenetrable. I thought they were wimps."

The idea that ecosystems dominated by pine, an invasive species, were so thick that his workers couldn't even walk through them went against a central assumption of ecology: that native forests will be the lushest. Millennia of co-evolution should have created an ecosystem in which almost every niche is filled, converting the available energy into trees and other species in the most efficient way. Conservationists also generally assume that native ecosystems

contribute best to ecosystem services.

Lugo went to see for himself. Sure enough, the pine plantations were bursting with vigour, far more so than nearby native-only forests of the same age. Lugo did a systematic study of the pine plantations and some mahogany ones, and found that the plantation understoreys were nearly as species rich, had greater above-ground biomass (the sheer weight of all the living things) and used nutrients more efficiently than the native forest understoreys. He submitted his results to the journal *Ecological Monographs*¹. Reviewers were horrified. In the end, it took almost a decade to get the paper past peer review.

Since then, Lugo has found many novel ecosystems in Puerto Rico and elsewhere that are much more diverse than native forests, but that are largely ignored by ecologists. "That diversity doesn't count because they are the wrong species," says Lugo, shaking his head. He's found alien trees that, by creating a shaded canopy on parched, degraded pastureland, make possible the establishment of native trees that could never cope with such an environment on their own. As a result he now finds it difficult to despise invasive trees as he thinks his colleagues do, and even embraces the change. "My parents and their parents saw one Puerto Rico," he says, "and I am going to see another Puerto Rico, and my children will see another."

Lugo wasn't the only researcher thinking along these lines, but it was not until 2006 that the new approach gained a manifesto — and a name. Lugo and 17 other researchers published a paper called "Novel ecosystems: theoretical and management aspects of the new ecological world order"² suggesting that such systems were



Joe Mascaro investigates biodiversity on Hawaii's Big Island.

GLOBAL COVERAGE

The amount of land taken up by novel ecosystems, defined as unused lands embedded within agricultural and urban landscapes.



E. ELLIS

worth scientific attention. To demonstrate the depth of resistance to the idea, the published paper quoted referees' comments on the submitted manuscript: "One reviewer commented that the examples are ecological disasters, where biodiversity has been decimated and ecosystem functions are in tatters, and that 'it is hard to make lemonade out of these lemons'." But Lugo and his colleagues saw it in a different light: "We are heading towards a situation where there are more lemons than lemonade," they wrote, "and we need to recognize this and determine what to do with the lemons."

Lemons can have their own value, says restoration ecologist Richard Hobbs, lead author of the paper and now at the University of Western Australia in Crawley. Some novel ecosystems, he says, are "alternative stable states", relatively entrenched ecosystems that would be very difficult to drag back to historical conditions.

Around the time the paper came out, Mascaro became interested in Lugo's work and set out to see if his results could be replicated on the windward side of Hawaii's Big Island. Were the many novel ecosystems on the islands nurturing any native species? Were they providing ecosystem services? He studied 46 forests growing on lava flows of varying ages at various altitudes and dominated by a variety of species, including *albizia* (*Falcariaria moluccana*), a fast-growing tree from southeast Asia, and Australian ironwood (*Casuarina equisetifolia*). He found that, on average, the forests had as many species as native forests. But by and large they weren't incubating natives as they seemed to in Puerto Rico³.

Part of the reason for the difference may lie in the uniqueness of Hawaiian flora, which evolved in isolation for up to 30 million years⁴. Not many plants got to Hawaii in the first place, so competition and predation pressures weren't very fierce. Without having to worry about being eaten by anything larger than an insect, raspberries and roses lost their thorns and mints lost their minty defence chemicals. When people introduced plants from other parts of the world, along with their attendant herbivores, Hawaiian plants couldn't compete.

Futuristic perspective

But Mascaro's results didn't put him off the novel-ecosystem concept. For one, he found that in many measures of forest productivity, such as nutrient cycling and biomass, novel forests matched or out-produced the native forests. They might not be 'natural' in the eyes of purists, but they are behaving exactly as they should. "These ecosystems, like it or not, are going to be driving most of the natural processes on Earth," he said at the 2008 Ecological Society of America meeting in Milwaukee, Wisconsin. It's a message that Peter Kareiva, chief scientist at the Nature Conservancy in Seattle, Washington, wants to see move from the academic world to the world of conservation management. "You hear conservationists talk about what they want to save, what they want to stop," he says. "They should

talk about what they want the world to look like in 50 years." Studies of novel ecosystems could help conservationists to "face the facts and be strategic", Kareiva says, rather than trying to beat back the unceasing tide of change.

Kareiva is a great fan of the ecosystem-services argument for preserving nature. But he admits that the problem of what to do when novel ecosystems provide better services than the native ones is "a question we don't talk about that much". Nevertheless, he is willing to imagine a world in which, for example, exotic strains of the reed *Phragmites* are allowed to thrive in US wetlands because they provide a

great habitat for birds, rather than be torn out in an expensive and potentially fruitless attempt to return native vegetation to dominance.

Ecosystem-service arguments are powerful enough to get some ecologists to abandon, or at least put to one side, their deep distrust of novel ecosystems. Like many of his peers, Shahid Naeem, an ecologist at Columbia University in New

York, says he "would love to get rid of every invasive species on the planet and put all the native species back in their place". Yet he's willing to see what can be made of novel ecosystems as he feels an imperative to improve conditions for the billions of humans on Earth.

The idea that novel ecosystems provide welcome diversity has also gained traction. Thinking on 'invasive species' has mellowed

"If we want to debate whether to conserve novel ecosystems, we will have to deal with the risk they pose to other systems."

— Joe Mascaro

significantly since the field was first established in the 1950s. Newer work by the likes of Mark Davis at Macalester College in Saint Paul, Minnesota, and Dov Sax at Brown University in Providence, Rhode Island, has shown that the vast majority of species that humans move around can slot into new ecosystems without driving anything else extinct, and that the common vision of invasive plants forming dense monocultural stands that take over everything else in their path is actually the exception. Yet the newcomers in novel systems can still be a genuine worry.

Peter Vitousek, an expert on Hawaiian biodiversity at Stanford University in California, would put albizia forests in the category of dangerous invaders, because they wipe out stands of the native 'ōhi'a tree (*Metrosideros polymorpha*). He acknowledges the services that novel ecosystems provide and that "they may even support native biological diversity in some important circumstances". But, he adds, "as with many good ideas, [tolerance of novel ecosystems] can be taken to an extreme at which it is no longer useful. I think most of the albizia-dominated stands of Hawaii represent that extreme." His point is well illustrated where one of Mascaro's albizia forests abuts a native 'ōhi'a forest. The albizia trees on the boundary actually lean out towards the 'ōhi'a — growing sideways to escape the shade of the next row in, encroaching on the natives' sunlight and looking poised to usurp them. It is a menacing spectacle, and an apt symbol for their tireless expansion.

Mascaro grants the point. "I can understand where a manager wants to bulldoze an albizia forest if they are worried that it is going to exterminate an ecosystem type that is the last on Earth," he says. "If we want to debate whether to use or conserve novel ecosystems, we will always have to deal with the risk they pose to other systems. But at the moment, we're scarcely debating it at all."

Novel ecosystems are likely to cause at least some extinctions. For example, species that have evolved dependent relationships with other species are less likely to do well in a world in which the pot is stirred and everything is redistributed. Hawaiian honeycreepers, beautiful birds that often feed only on one type of flower, are not doing well; several are already extinct. So for those who care about slowing or stopping the rate of such extinctions, novel ecosystems are a net negative.

James Gibbs, an ecologist at the State University of New York in Syracuse, subscribes to this view. "I think celebrating [novel ecosystems] as equivalent or improved is not appropriate." As an example, he points to Clear Lake in Northern California, where the number of fish species has risen from 12 to 25 since 1800.

Sounds like a success story. But, says Gibbs, species that were found only in that lake were replaced with fish that are common elsewhere — so there was a net loss in biodiversity. A similar caveat may hold for the genetic diversity hidden within a species. Forests dominated by the offspring of a handful of exotic colonizers could be less genetically diverse than forests that have sat there for thousands of years.

A question of values

In the end, the question of novel ecosystems, like so many questions in ecology and conservation, boils down to what should be valued most in nature. For people who value processes, such as Mascaro, novel ecosystems are great hubs of active evolution. For those who value ecosystem services, any novel ecosystem could be better or worse than what came before depending on how it operates. For those who care about global extinctions or about preserving historical ecosystems, they are bad news. Gibbs says he values the exquisite complexity of ecosystems that have evolved together over thousands or millions of years. "Why are we worried about the extinction of languages, the roots of music, all these weird cuisines?" he asks. "There is something about diversity and our need to steward it. It is the subtlety and the nuance and complexity that makes life interesting." Novel ecosystems seem, to him, to lack this value, to be samey and artificial, "sort of like eating at McDonalds".

To Kareiva, though, that attitude is "one of the reasons the conservation movement is failing. To think there is some kind of garden of Eden pristine ecosystem. There is none! That

view is just going to get us nowhere."

Indeed, the Garden of Eden view, in which ecosystems are static, is no longer widely held. This means that novel ecosystems, far from being a new phenomenon, simply represent the latest changes on a dynamic Earth. Gradual climatic changes and sheer randomness mean that some species wander around continents over vast timescales, fleeing glaciers, splitting up and reforming. This is why Davis and some others do not like the 'novel' label. "Ecosystems are always new, from one year to the next," says Davis. "Ecosystems are always encountering

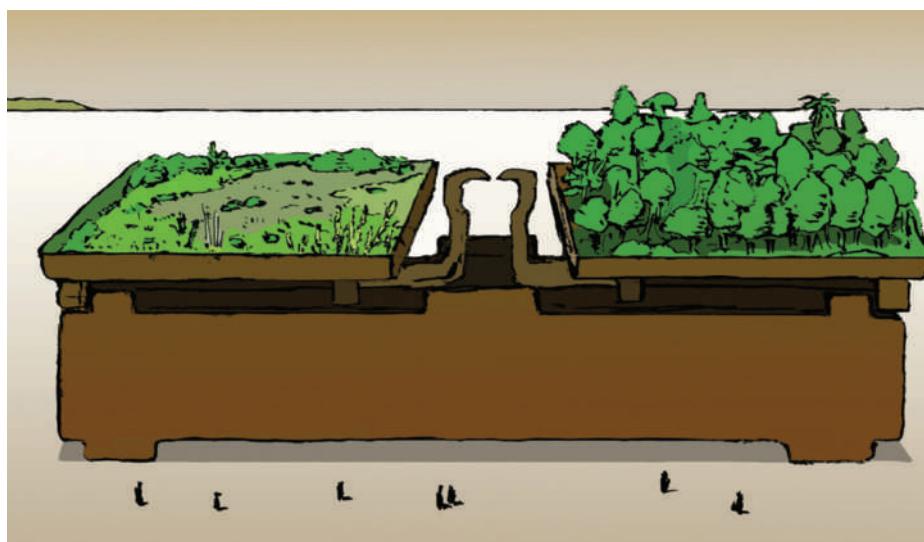
new species — it might be not from another country but from 100 metres upstream. Much more accurate would be to refer to these as 'rapidly changing' ecosystems — but I guess that is not catchy enough."

Standing in his Hawaiian forest, Mascaro is all too aware of change — and it is something he values, even if humans did have a hand in the process. He never swore allegiance to preserving ecosystems as they were before humans arrived, as many conservationists of an older generation did. "People come up to me and say 'it sounds like you've given up,'" says Mascaro. "I want to say 'I never took up arms, my man.' This isn't about conceding defeat; it is about a new approach." ■

Emma Marris writes for Nature from Columbia, Missouri.

1. Lugo, A. *Ecol. Monogr.* **62**, 2–41 (1992).
2. Hobbs, R. J. et al. *Global Ecol. Biogeogr.* **15**, 1–7 (2006).
3. Mascaro, J., Becklund, K. K., Hughes, R. F. & Schnitzer, S. A. *Forest Ecol. Manage.* **256**, 593–606 (2008).
4. Ziegler, A. *Hawaiian Natural History, Ecology and Evolution* 157 (University of Hawaii Press, 2002).

See Editorial, page 435.





Paul Octavious

The globe's green avenger

Maurice Strong has shaped how nations respond to planetary crises. **Ehsan Masood** meets the man whose successes — and failures — laid the groundwork for the current climate talks.

In 1971, Maurice Strong did something most people would find unthinkable: he showed up at the office of India's Prime Minister Indira Gandhi to tell her that she was wrong. The formidable Gandhi, like the leaders of other developing nations, wanted to boycott the United Nations' first major international environmental conference, planned for the following year in Stockholm. The developing nations feared that environmental agreements would undermine their ability to prosper — a concern that still resonates today.

As the leader of the upcoming conference, Strong regarded the developing nations as essential players. So, against the advice of Indian diplomats, he flew to New Delhi and urged Gandhi to drop the boycott and attend the conference as an advocate for the developing world.

"She had a habit of falling silent during conversations and said absolutely nothing for 10 or 12 minutes," Strong recalls. "Now, I've lived among the Inuit, and they, too, fall silent during conversations, although for much longer



"Maurice had this ability to reach out and grab world leaders by the arm, whether they liked it or not."

— Jonathan Lash

periods, so this didn't bother me at all. I waited for her reply and when she finally spoke, she said 'yes.'

Strong has spent much of his life trying to make the world see matters his way, sometimes by remaining silent but more often through a mixture of negotiation and persuasion. He helped to found the United Nations Environment Programme (UNEP) in 1972. He chaired the 1992 Earth Summit in Rio de Janeiro, Brazil, that produced the first climate treaty. And he mentored a generation of environmental, political and business leaders around the world.

Many of them gathered in Gland, Switzerland, this month at the headquarters of the International Union for the Conservation of Nature (IUCN) to celebrate Strong's 80th birthday. It was also an occasion to take stock of the planet and discuss the round

of climate negotiations that will conclude in December in Copenhagen. The chances for success in Denmark depend to a large degree on what Strong has accomplished over the past 40 years, and on what he was unable to achieve.

From the beginning, Strong had outsized ambition. Born into poverty in the Canadian town of Oak Lake, Manitoba, at the start of the Great Depression, he writes in his autobiography *Where on Earth are We Going?* that his childhood dream was to devote his life to the protection of nature and to work for world peace, having lived through the Second World War and seen its effects on humans and on the environment. When the war ended and countries were discussing the establishment of the UN, Strong says, "I wrote in my diary: 'this is where I want to be.'"

Ascent of nations

Strong started at the bottom, working in the office that issues security passes at the UN building in New York. Later he would rise, via posts in the Canadian government and in business, to chair the Stockholm and Rio conferences on the environment. He helped to set up Canada's national oil company Petro Canada as well as its International Development Research Centre, and he led the team that proposed a plan to reform the UN. Strong's childhood instilled a determination never again to be poor, and he grew rich leading various companies and through property investment.

In many ways, Strong's greatest legacy is

the creation of UNEP, which emerged out of the 1972 Stockholm conference and has since played a catalytic part in international environmental issues. The organization oversaw the process that created the Montreal Protocol of 1987, a treaty to phase-out ozone-depleting compounds. In 1988, UNEP joined with the World Meteorological Organization to create the Intergovernmental Panel on Climate Change. Four years later in Rio, UNEP and the environmental movement shaped new international agreements on climate change and biodiversity, and laid the groundwork for a treaty to combat desertification. The climate convention provided the road map for the Kyoto Protocol — in which nations pledged to cut greenhouse-gas emissions — and established the process for the climate talks.

Bridge-building

For the Stockholm and Rio conferences to succeed, Strong knew he needed to earn the trust of developing countries. But he also built bridges with the scientific and business communities and he opened up both events to participation from non-governmental organizations — often against the wishes of governments.

"Maurice really pushed the door for public engagement," says James Gustave Speth, a former administrator of the UN Development Programme and founder of the World Resources Institute, an environmental think-tank in Washington DC. "You have to remember that this had never happened before and it had ramifications across all other UN processes."

Yet despite Strong's successes in bringing people and nations together, the policy architecture he created has done little so far to make the world safer. Aside from the successful global effort to protect the ozone layer, international environmental agreements have faltered. In particular, they lacked mechanisms to verify whether countries met their commitments or to penalize those that fell short. This led to the anticipated failure of the Kyoto treaty, and will also doom the parallel commitment to slow down the rate of loss of biodiversity by 2010. The consensus at the meeting in Switzerland was that the fault lies with the current environmental policy structure.

Strong acknowledges that the treaties have suffered because they allowed "governments to say one thing, but do the opposite". The fault may lie with the expectations that Strong and others had. The entire environmental movement had a misplaced confidence in the ability of governments to deliver on their promises, says Speth.

That realization has tempered hopes for the climate negotiations this year, and Strong doubts whether December's Copenhagen climate summit will succeed. "Radical thinking is needed," he says, "but there is no evidence that governments are ready for that." There is no point to an agreement, he says, unless it has "binding real penalties, fines and trade bans that are designed to make agreements enforceable", rather like what happens in the World Trade Organization or the International Atomic Energy Agency.

Another problem with past environmental agreements is the relationship between the richer and poorer countries. Developing nations have long insisted that if they were going to be forced to grow sustainably — unlike the richer nations — then substantial financial support should be made available. But the wealthy nations baulked at that request, in part because they didn't want to be seen as accepting responsibility for global environmental problems. It has taken nearly four decades for a potential solution to emerge. UK Prime Minister Gordon Brown proposed last month to start a climate fund that would give US\$100-billion to developing countries each year, paid for by the world's wealthiest nations.

A third issue that has dogged Strong is his poor relations with US administrations — and his belief that global agreements can work without the United States. "I did spend a lot of time in America," he says. "But America already had many people who could speak on its behalf. Yes, I could have put more time in, but even the Americans were telling me 'you've got to get the developing countries on side'. I used to

"There is no evidence that governments are ready for radical thinking."

— Maurice Strong

tell the Americans that if I am seen to be in your pocket, then I am of no use to anyone."

To this day, Strong gets bad press in the United States, where reporters often focus on potential conflicts of interest that affected him while he was working in the upper tiers of the UN.

Little black book

Yet Strong has no shortage of supporters, especially among the world leaders he helped to groom. He has a legendary record in spotting promising individuals, which is evident in the roll-call of people whom he hired early in their careers. Canada's former prime minister Paul Martin, James Wolfensohn, who led the World Bank, Yolanda Kakabadse, former environment minister of Ecuador and now president of the environmental group WWF, and many serving heads of UN agencies are Strong protégés and cite him as a major influence in their lives.

When the IUCN was having financial problems in the 1980s, "Maurice stepped in and came to our rescue", says Julia Marton-Lefèvre, IUCN director-general. "He would make the phone calls or get us a seat at the right table," she says. "Maurice had this ability to reach out and grab world leaders by the arm, whether they liked it or not," says Jonathan Lash, president of the World Resources Institute. "He used his access and networks shamelessly for all of us. Long before iPhones, the net and Twitter, we had Maurice."

At the start of the Rio conference, Strong had insisted (and as usual got his way) on a private meeting between himself, the UN secretary-general and more than 100 heads of government, without their minders and attendants. In his address to them, Strong said that the planet had only one real chance, and that they stood between success and failure.

Strong was wrong about the one chance. The Rio climate treaty, and its progeny, did not stop global warming, and there is debate about whether they have reduced the problem in any substantial way. But politics and the climate have evolved since Strong urged leaders towards action. The world has yet another chance in Copenhagen and the outcome depends largely on whether countries have learned from their past mistakes.

Ehsan Masood teaches international science policy at Imperial College London.



Maurice Strong (left, with former Soviet leader Mikhail Gorbachev) has been able to bring world leaders together.

CORRESPONDENCE

Outrage at high price paid for a fossil

SIR — In your Editorial 'Media frenzy' (*Nature* **459**, 484; 2009), you discuss the hype surrounding the description of a 47-million-year-old fossil primate (J. L. Franzen *et al.* *PLoS ONE* **4**, e5723; 2009). This fossil was publicized worldwide by the press using banners such as "the eighth wonder of the world" and "the missing link in human evolution", although no such claims were made by the authors in the paper. But even more outrageous, in our view, was the earlier sale of this specimen for a huge sum.

The fossil, nicknamed 'Ida', was allegedly found by an amateur collector (C. Tudge *The Link* 68–69; Little, Brown, 2009) in 1983 at the Grube Messel site in Germany.

Ida's collector separated this fossil into two parts, which was unfortunate because of the scientific value of completeness. The less-complete part, now known as the B side, reportedly had some of its anatomical features fabricated to make it seem more complete (see J. L. Franzen *et al.* *PLoS ONE* **4**, e5723; 2009), perhaps to facilitate its independent sale.

The B side was sold to Burkhard Pohl of the Wyoming Dinosaur Center in Thermopolis, then sent to Jens Franzen of the Senckenberg Research Institute in Frankfurt, Germany, who described it in 1994. Meanwhile, the more complete part of Ida, now known as side A, was embedded in resin and framed. It is believed to have been kept in the anonymous collector's home for 23 years until 2006, when its photograph was shown to Jørn Hurum of the Natural History Museum in Oslo.

The asking price for side A was reportedly US\$1 million, although Hurum is alleged to have eventually paid about \$750,000 for it (see *The Times Online*, 28 May 2009). The publicity barrage surrounding this fossil seems



to have been amplified by its purchase for this sum of money.

In our view, such objectionable pricing and publicity can only increase the difficulty of scientific collecting by encouraging the commercial exploitation of sites and the disappearance of fossils into private collections. We believe that payments on this scale are detrimental to scientific investigation, and respectable institutions should not be responsible for making or publicizing them. We strongly believe that fossils should not have any commercial value.

Elwyn L. Simons, Friderun Ankel-Simons, Prithijit S. Chatrath Duke Lemur Center, Durham, North Carolina
e-mail: esimons@duke.edu

Richard S. Kay, Blythe Williams
Evolutionary Anthropology
Department, Duke University
John G. Fleagle Department of Anatomy, State University of New York, Long Island

Daniel L. Gebo Department of Anthropology, Northern Illinois University

Christopher K. Beard, Mary Dawson
Carnegie Museum, Pittsburgh

Ian Tattersall Department of Anthropology, American Museum of Natural History

Kenneth D. Rose Center for Functional Anatomy and Evolution, Johns Hopkins University School of Medicine

future prosperity and in tackling global challenges. It has also never been more important to justify how we spend public money. This is why the research councils are working together to highlight the impact of the research we fund.

Some people have reacted strongly to the requirement for proposals to include an outline of the "potential economic impact" of the research. This has been taken to mean purely financial or commercial impact, whereas the definition consistently used by the research councils is much broader and embraces all the diverse ways in which research-related knowledge and skills could benefit individuals, organizations and nations. For example, high-quality research can lead to improved environmental measures, better communications, new products and services and better-informed public policy.

The primary criteria used to judge proposals remain international excellence and quality. We simply want to ensure that our researchers have considered the impact of their work from the outset.

There have been concerns that the emphasis on impact will draw attention away from blue-skies research. But the EPSRC is fully committed to investigator-led research and understands that this is where future generations of technology will come from.

Our mission remains unchanged — to promote and support, by any means, high-quality basic, strategic and applied research, thereby contributing to the economic competitiveness of the United Kingdom and our quality of life.

We must all work together to demonstrate the impact that science and engineering have on society, to ensure continued public support and government funding for research.

David Delpy Engineering and Physical Sciences Research Council, Polaris House, North Star Avenue, Swindon SN2 1ET, UK
e-mail david.delpy@epsrc.ac.uk

"Smaller investments reduce the pressure on companies and allow them to become more flexible in their business strategies." John Browning, page 459

Time running out to deal with banks of greenhouse gases

SIR — In your Special Report 'Cutting out the chemicals', you discuss the possible shift of regulatory control of hydrofluorocarbons (HFCs) from the UN Framework Convention on Climate Change to the Montreal Protocol (*Nature* **457**, 518–519; 2009). Since then, amendments to the protocol have been proposed that would establish HFC phase-down schedules, in parallel with similar provisions in US climate legislation — as mentioned in your News story 'Climate burden of refrigerants rockets' (*Nature* **459**, 1040–1041; 2009). However, although a bridge may be under construction to address the problem of HFCs, there is another gap that requires action.

Neither treaty controls emissions of the remaining chlorofluorocarbons and other ozone-depleting substances that HFCs were designed to replace. Besides the damage they cause to stratospheric ozone, these are also greenhouse gases that are up to 11,000 times more potent than carbon dioxide. Large quantities of chlorofluorocarbons produced before the phase-out deadlines remain in use in older appliances and buildings. These 'banks' are capable of releasing into the atmosphere the equivalent of 18 gigatonnes of carbon dioxide (see the International Panel on Climate Change and the Technology and Economic Assessment Panel's *Safeguarding the Ozone Layer and the Global Climate System*, 2005, at <http://tinyurl.com/ndtd9z>), about one-third of which will be emitted by 2015 unless recovered and destroyed under proper incentives.

Policy-makers have an immediate opportunity to prevent hundreds of millions of tonnes of greenhouse-gas emissions every year. Possible strategies could include: counting banked ozone-depleting substances as

controlled greenhouse gases; allowing greenhouse-gas offsets for certified destruction of banked ozone-depleting substances; setting rigorous standards and protocols to quantify and verify projects for destruction of ozone-depleting substances; and establishing incentives for replacement technologies that avoid substitution of ozone-depleting substances with other greenhouse gases.

Recognizing these banks as greenhouse gases in domestic legislation and international agreements would spur carbon markets to finance cost-effective collection, as well as transport and destruction of unwanted ozone-depleting substances. In the interim, regional and state cap-and-trade schemes could take comparable steps to mobilize projects, permanently removing a major threat to ozone and climate, and accelerate the transition to advanced technologies. There is only a narrow window to address this opportunity, but the timing is right for action.

Jeff Cohen EOS Climate, San Francisco, California, USA
e-mail: jcohen@eosclimate.com

Petitioning for a revised statement on climate change

SIR — We write in response to your issue discussing "the coming climate crunch", including the Editorial 'Time to act' (*Nature* **459**, 1077–1078; 2009). We feel it is alarmist.

We are among more than 50 current and former members of the American Physical Society (APS) who have signed an open letter to the APS Council this month, calling for a reconsideration of its November 2007 policy statement on climate change (see open letter at <http://tinyurl.com/lg266u>; APS statement at <http://tinyurl.com/56zqxr>). The letter proposes an alternative statement, which the signatories believe to be a

more accurate representation of the current scientific evidence. It requests that an objective scientific process be established, devoid of political or financial agendas, to help prevent subversion of the scientific process and the intolerance towards scientific disagreement that pervades the climate issue.

On 1 May 2009, the APS Council decided to review its current statement via a high-level subcommittee of respected senior scientists. We applaud this decision. It is the first such reappraisal by a major scientific professional society that we are aware of, and we hope it will lead to meaningful change that reflects a more balanced view of climate-change issues.

S. Fred Singer University of Virginia
Hal Lewis University of California, Santa Barbara
Will Happer Princeton University
Larry Gould University of Hartford
Roger Cohen Durango, Colorado
Robert H. Austin Princeton University, Princeton, New Jersey 08544, USA
e-mail: austin@princeton.edu

Indigenous people defend rainforest as well as their rights

SIR — What can we learn about the relationship between native peoples and environmentalism in the Amazon from last month's tragic clashes between indigenous protesters and government security forces in Bagua, Peru? The event, in which more than 30 people were killed, underscores the overlooked and sometimes hazardous role of indigenous groups and organizations in confronting powerful commercial interests in the Amazon.

The assumption that heavily forested, indigenous-occupied lands are unproductive has fuelled colonization, deforestation and tribal displacement in Amazonian Peru for a century. Following enactment of the United States–Peru Trade Promotion Agreement in February this year, Peruvian

president Alan García issued a series of decrees aimed at opening native communities to oil, mining, logging and other economic ventures. Indigenous Awajun and Wampis protesters, calling for the repeal of these decrees, set up a roadblock in April that was violently dispersed on 5 June by security forces. After two weeks of intense international media attention, the government repealed a few of the decrees.

The incident, which has been called the 'Bagua massacre', provides a sobering lesson for the conservation community. The 'ecologically noble savage' debate has revolved around whether demographically growing and politically engaged indigenous populations represent a threat to biodiversity conservation or an opportunity for it. Defenders of native rights to vast tracts of Amazonia say that indigenous reserves deter deforestation, whereas some conservationists say that their exploitation of natural resources has adverse effects (see S. Schwartzman *et al.* *Conserv. Biol.* **14**, 1351–1357, and J. Terborgh *Conserv. Biol.* **14**, 1358–1361; 2000). But ultimately, the dual goals of nature conservation and indigenous rights have proved compatible. Three-quarters of Latin American national parks include native inhabitants, and Amazonian indigenous reserves maintain forest cover and refuges for many animal species — unlike surrounding colonized areas, which suffer heavy losses to biodiversity and forest cover.

Indigenous people's intimate connection with the Amazon rainforest has led them, countless times, to lay down their lives in defence of their territories and, incidentally or not, the biodiversity and carbon stocks they harbour. Hard-core conservationists should remember these sacrifices when factoring out the 'ecologically noble savage'.

Glenn H. Shepard Department of Anthropology, Museu Paraense Emílio Goeldi, Av. Perimetral, 1901 Terra Firme, Belém, Pará 66077–830, Brazil
e-mail: gshepadjr@gmail.com

ESSAY

The incredible shrinking venture capital

Venture funding is declining quickly and is unlikely to bounce back. But less money means lower expectations — good news for smaller science start-ups, says **John Browning**.

The days when stock options put Ferraris into the car parks of start-ups are over, and venture capital is drying up. In the first quarter of 2009, US venture-capital investments in the life sciences fell 40% from the fourth quarter of 2008, in biotechnology they fell 46%, and in software 42%. Even investment in green technologies — such as those designed to improve energy efficiency — was down 84%, more than wiping out the 54% increase from 2007 to 2008.

This isn't necessarily all bad news. The venture-capital boom of the 1990s created expectations that many companies cannot meet. As the amount of venture capital shrinks, both venture capitalists and entrepreneurs are rethinking some of their basic assumptions. Having less money might make companies healthier, given the reduced pressure to be a skyrocketing success and the increased flexibility for business strategy and growth. Although the conventional venture-capital model has best suited fast-growing IT or biotech companies, it could begin to suit science and technology start-ups aimed at smaller markets.

Venture capitalists usually look for companies that will grow quickly — as a rule of thumb, with sales of at least US\$50 million within six years — and that can then be sold, either to a larger firm or on to the stock market. Without a sale, venture capitalists can't return profits to those who invested the money in the first place. The microelectronics and Internet booms of the 1990s and early 2000s provided both fast growth and sales; the lure of a Google-like financial cornucopia inspired a fivefold increase in the amount of capital in venture funds between 1996 and 2001. But more and bigger sales were needed to keep investors happy.

Risky business

This style of investing can build great companies. Think of Google, Intel, Apple, Cisco Systems and Genentech. Mercenary though it may be, the focus of creating a sellable company forces managers to make the most of their opportunities. Venture capitalists are also tolerant of risk: a typical fund would expect about a third of its investments to fail outright. But what venture capitalists can't tolerate is the merely average. They have a term for companies that reach, say, \$10 million in sales and

neither lose so much money that they go out of business, nor grab the markets to continue growing and fetch a high sale price. They call them the 'living dead'. Where many might see a nice little company — including biotech ventures whose names would be familiar to many — a venture capitalist sees mostly frustration. Today, with capital markets crunched and sales slow, more and more investments look

Conway, recently financed her travel website by selling cupcakes.

Others have been looking harder at 'angel financing', in which wealthy individuals put their own money into deals. These individuals still want a return on their investment, but they are more willing to take it as a stream of dividends from a company that has found a valuable niche, rather than as profit from the sale of

the company. Although the amount of angel financing in the United States shrank by about 26% from 2007 to 2008, it was still about \$19 billion for 2008, not far below traditional venture capital, according to a study by the Center for Venture Research, at the University of New Hampshire in Durham. What is more promising is that the money was spread across roughly the same number of deals in 2008 as in 2007: about 55,000.

Given the lacklustre returns of traditional investment strategies, venture capitalists are also looking to do more with less. Marc Andreessen, co-founder of Netscape Communications and a pillar of the Silicon Valley establishment, recently co-launched a venture firm that plans to invest as little as \$50,000 per start-up — far less than the \$3 million considered to be a minimum by many venture capitalists. Although it is early days, efforts such as this might reshape venture capitalism. Without the weight of Googlesque expectations on their shoulders, companies that might have joined the ranks of the living dead could start to look lively. A start-up focused on a non-blockbuster drug or diagnostic test might now find itself with an attractive niche market, garnering the attention of venture capitalists who would usually have avoided this type of limited-growth company.

Smaller investments will force entrepreneurs to work harder — no more plush offices or fridges stocked with designer fruit juice. But, because the returns demanded by investors are proportional to the amounts put in, smaller investments also reduce the pressure on companies and allow them to become more flexible in their business strategies. And that is what entrepreneurship needs most.

John Browning is a writer and consultant. He co-founded the First Tuesday networking forums for venture capitalists and innovators. e-mail: jb@poyais.com



as though they fall into this category.

Returns on venture-capital investments have sagged steadily since the Internet boom years. Now both venture capitalists and their investors are losing patience. After a 7% decline in 2007, the collective assets of US venture capitalists shrank by 24% in 2008, to \$197 billion. The result was fewer investments, falling from \$8 billion across 1,058 deals in the fourth quarter of 2007 to \$3 billion for 549 deals in the first quarter of 2009 — the lowest amount of investment since 1997. Paul Kedrosky, a former venture capitalist turned industry commentator, predicts the money isn't coming back. He estimates that the venture-capital industry will settle down to about half the size it was in the early 2000s, or about the same as it was in the late 1990s.

In the face of such declines, some are eschewing outside investment altogether and financing companies with their own resources. In IT, the cost of starting a business has fallen dramatically; one US entrepreneur, Paula

BOOKS & ARTS

Great inventions of life

A book setting out the ten greatest transformations delivered by evolution contains surprises but neglects crucial innovations such as proteins and embryos, **Lewis Wolpert** finds.

Life Ascending: The Ten Great Inventions of Evolution

by Nick Lane

W. W. Norton & Co./Profile Books: 2009.

352 pp/288 pp. \$26.95/£18.99

The variety of living things is astonishing — as is their evolution. Nick Lane describes in *Life Ascending* “the greatest inventions of evolution, how each one transformed the living world, and how we humans have learned to read this past with an ingenuity that rivals nature herself”. The ten inventions he has chosen are the origin of life, DNA, photosynthesis, the complex cell, sex, movement, sight, hot blood, consciousness and death. Not everyone will agree with his choice. But his writing and explanations are excellent and imaginative and, similar to life itself, the book is full of surprises.

The origin of life — the evolution of the single cell from which all life is derived — is not yet solved. Particularly difficult to explain is the evolution of cell division. The first cells were simple bacteria, but they were complex in that their constituent molecules interacted in a reliable way. Eukaryotic cells with a nucleus and with mitochondria came later. The book mentions the origin of mitochondria — thought to have occurred when a primitive eukaryotic cell swallowed a bacterium — but explaining the origin of the nucleus is harder.

Surprisingly, Lane pays little attention to proteins. He discusses genes in detail, but in the cell they are passive, coding for the true workers: the proteins. Proteins provide the cell with the ability to synthesize and break down molecules, and they turn genes on and off and replicate DNA. They give the cell its basic structure. But how did proteins evolve?

A major omission from Lane's list of inventions is the embryo and the development of complex forms from a single cell. The origin of multicellularity is not discussed, yet it is fundamental to the evolution of all animals. My view is that it may have evolved when a group of cells stuck together after dividing and then, when food was in short supply, some cells survived by eating each other. This could have given rise to the egg, which is a cell fed by other cells.

Lane believes that it was the evolution of movement after the Permian extinction some 250 million years ago that transfigured life. Flowering plants could evolve once there were motile organisms to fertilize them. A key stage



Flowering plants that rely on pollinators could evolve once organisms had developed the ability to move.

was the origin of muscle proteins, which are not present in bacteria. But Lane neglects to relate this development to the origin of the brain, a major function of which is to control movement.

Looking at consciousness, Lane admits that there is as yet no understanding of how neuron firing gives rise to intense personal sensations. He pays little attention to language and to what makes us human. In my view, it is our belief in physical causality that led to tool-making and drove human evolution.

The evolution of vision is fundamental — 95% of animal species have eyes, and the ability to see may have played a key part in the rapid burst of evolution known as the Cambrian explosion as prey had to adopt new survival strategies. Lane counters the critics who claim that an eye must be fully functioning to be adaptive: the naked retina of the supposedly blind shrimp is indeed adaptive. A surprising example of the cleverness of evolution is the nature of the crystallins that make up most of the proteins in the human lens. Many other crystallins are enzymes with housekeeping functions in the body that are unrelated to vision.

On evolution and sex, Lane argues that “men are a heavy cost” and states that “a woman who solved the problem of virgin birth would be a worthy madonna”. He regards as a great advantage of sex that it allows good genes to recombine away from the junk residing in their genetic background.

On the evolution of warm blood, which lets animals such as birds and mammals maintain a constant temperature, Lane points out that it is a surprising outcome given the energy expended. He explains it as a means of giving stamina to animals.

Death, in Lane's view, is about ageing: “Death may seem a cruel cosmic joke, but ageing is mirthless.” Only the germ line is immortal; he supports the ‘disposable soma’ idea that evolution has no interest in animals once they have reproduced.

Life Ascending is a fascinating book for anyone interested in life and evolution, and how these discoveries were made. ■

Lewis Wolpert is emeritus professor of cell and developmental biology at University College London, Gower Street, London WC1E 6BT, UK.
e-mail: l.wolpert@ucl.ac.uk

Digesting evolution

The Evolution of Obesity

by Michael L. Power and Jay Schulkin
Johns Hopkins University Press: 2009.
408 pp. \$40, £21

The Evolution of Obesity is a fusion cuisine that brings together proximate explanations of weight-regulation mechanisms and the evolutionary reasons for why these can fail. The result is a rich meal that is memorable, if slightly hard to digest. The book is one of the first to use an evolutionary framework to analyse a major body of neuroendocrine knowledge about a specific condition.

By considering the adaptive functions of the physiological mechanisms of weight regulation, Michael Power and Jay Schulkin construct this framework to try to make sense of a huge amount of disconnected knowledge. They are in the vanguard of those who recognize a disturbing fundamental principle: natural selection shapes systems that are nearly indescribably complex. After millions of years of selection and genetic drift, the mechanisms responsible for a particular biological function are so intertwined and multifunctional that no simple description can be accurate.

Attributing one function to a hormone is attractive, but often wrong. For example, the strong correlation of leptin levels with body-fat mass seems to suggest that leptin exists mainly to limit body weight. But its role is not so clear: administering it to people who are obese does not reduce appetite or weight. Low leptin levels signal the need to seek food, but leptin is only one part of the complex system that regulates eating. Power and Schulkin provide the real story: leptin's functions depend "on the tissue, the state of the organism, especially the expression of other information molecules such as insulin, glucocorticoids, and CRH, and also on the age of the organism. Leptin has many functions, which change with time, tissue, and circumstances."

The authors come to similar conclusions for dozens of information molecules that are involved in feeding regulation — including the hormones ghrelin, corticotropin-releasing hormone and cholecystokinins. Although their conclusion about the complexity of systems shaped by selection will distress readers who want a simple story, it is profound beyond the realm of weight regulation. Power and Schulkin argue that a full understanding will come only from "integrating the reductionist methodology into a more all-inclusive approach".

The book covers nearly all the main issues in

obesity research. Geneticist James Neel's theory of the 'thrifty genotype' — shaped to store fat to cope with food scarcity — receives a full critique. Similarly detailed is the authors' treatment of new suggestions that thrifty phenotypes might be induced in developing embryos whose calorie intake is limited. They discuss adipose tissue, where fat is stored, as an endocrine organ that secretes at least ten information molecules, and describe the role of cytokines as mediators of the tissue damage associated with abdominal obesity.

The epidemiology of the recent obesity epidemic — manifesting as a threefold increase in obesity in the United States in just 50 years — is covered well. In an insightful section, Power and Schulkin ask why humans eat meals instead of feeding continuously. They eruditely connect explanations from behavioural ecology with the brain mechanisms responsible and with the social functions of sharing meals. Changes in diet since the Palaeolithic era are reviewed carefully, as are the interactions between meat eating and brain evolution.

"Attributing one function to a hormone is attractive, but often wrong."

Most books on obesity are heavy on opinion and light on facts. *The Evolution of Obesity* is the reverse. I wanted to hear the authors' considered views on the relative contributions of fructose-based beverages, exercise and other factors to the obesity epidemic. They might also have discussed in more depth how mechanisms that limit body weight are shaped by selection forces. The authors mention the increased rates of predation for heavy, slow individuals, but they do not analyse other trade-offs, such as

the risks associated with increased foraging time, the caloric costs of maintaining extra weight or the costs of those calories not being available to kin.

The ingredients in *The Evolution of Obesity* are somewhat unbalanced, favouring proximate over evolutionary causes. Yet the book goes far beyond anything else that is available on obesity. Power and Schulkin deserve much credit for their bold attempt to combine evolutionary and reductionist explanations, and for their unflinching acknowledgement of complexity. ■

Randolph M. Nesse is professor of psychiatry and of psychology at the University of Michigan, Ann Arbor, Michigan 48109-1043, USA.
e-mail: nesse@umich.edu

Mathematical memories

The Housekeeper and the Professor

by Yoko Ogawa. Translated by Stephen Snyder
Harvill Secker/Picador: 2009. 192 pp.
£11.99/\$14

The Housekeeper and the Professor does for number theory what Jostein Gaarder's best-seller *Sophie's World* (Aschehoug, 1991) did for the history of philosophy, but with a far lighter touch. The narrator, ignorant of mathematics, becomes a surrogate for the average reader as the recipient of a great deal of detailed information. It is indisputably a novel, but it is unapologetically educational.

The Professor is an elderly mathematician whose memory lasts only 80 minutes, as a result of a head injury in a car accident in 1975. Everything that happened before his injury, including his vast grasp of number theory, remains in his mind, but every 80 minutes he has to start all over again.

With his amnesia being understandably hard on the people around him, the Professor is stashed in a cottage on his sister-in-law's

property. He has gone through nine frustrated housekeepers by the time the eponymous one is called into service. The Professor's suit is blanketed with hand-written notes to help him navigate the world, the most crucial of which says "My memory lasts only 80 minutes." Soon, a sketch of the new housekeeper joins the hundreds of other attached scraps, proving useful when she has to introduce herself every morning. Within minutes of their first meeting, the Professor starts to share his love of mathematics with her, and then with her son who joins them after school.

The Housekeeper, who like the Professor remains unnamed throughout the novel, is no ordinary servant. Possessed of intelligence, a keen curiosity and tremendous empathy, she is soon fascinated by her charge and surprisingly receptive to the Professor's lessons, as is her son. It helps that the Professor is an eloquent and skilled teacher. In one memorable scene, he likens the search for very large prime numbers to a quest through the desert wastelands. The Housekeeper has a vivid imagination and personifies the concepts: primes, in resisting division by any number other than

zero and one, are “stubborn”; the ‘abundant’ number 18, whose divisors’ sum is greater than itself, “secretly carried a heavy burden”; the ‘deficient’ number 14, whose divisors’ sum is less, “fell mute in the face of its terrible lack”. These tactics help readers to understand the concepts in a creative way, although not everyone will want to follow the strings of equations and diagrams that occasionally invade the prose.

Mathematics inspires a reverence in the Professor. There are repeated references to numbers as mystical entities, existing on the pages of “God’s notebook”. And in a world that is constantly being whisked out from underfoot, the Professor is comforted by the predictability of things he can remember:

mathematics, of course, but also statistics on baseball cards, or how to predict the position of Venus in the evening sky. A single mother with a dead-end job, the Housekeeper also finds maths a source of clarity: “I needed the sense that this invisible world was somehow propping up the visible one, that this one, true line extended infinitely, without width or area, confidently piercing through the shadows. Somehow, this line would help me find peace.”

The mathematical lessons weave through a narrative that is uneventful but pervaded



by a sense of beauty. Although the translation is fresh, and the characters could be of any nationality, there is a Japanese ambience. The language is as precise and graceful as a tea ceremony, and the backstory as sparsely sketched as *shodo* calligraphy. And if the Housekeeper’s newly kindled enthusiasm for numbers sometimes stretches the reader’s credulity, the book is so charming that the author is forgiven. ■

Jennifer Rohn is a cell biologist at University College London, London WC1E 6BT, UK, editor of www.LabLit.com and author of the novel *Experimental Heart*.
e-mail: jenny@lablit.com

HARVILL SECKER

Reflecting the impossible

Virtual Worlds: M. C. Escher and Paradox

Portland Art Museum, Portland, Oregon
Until 13 September

Two artists embody the saying that mathematics and art are so far apart they are practically neighbours — Leonardo da Vinci (1452–1519) and Maurits Cornelius Escher (1898–1972). Whereas Da Vinci searched for the possible, generating functional designs such as his flying machines, Escher searched for the impossible, creating images by distorting nature’s

rules. Escher’s techniques are explored in the retrospective exhibition *Virtual Worlds: M. C. Escher and Paradox*, now showing at the Portland Art Museum in Oregon.

Escher’s prints of tessellations form the heart of the show. *Heaven and Hell* (1960; pictured) is the most intricate, incorporating patterns that repeat at many size scales, inspired by the tile designs of the Alhambra palace in Granada, Spain. The print’s alternative name, *Circle Limit IV*, reflects the mathematical challenge that the artist undertook to make it.

To achieve visual balance, Escher insisted that the shrinking patterns converge towards a circular boundary. The patterns emerge “like rockets”, in his words, and flow along curved trajectories until they “lose themselves” again at the boundary. Constructing this sequence required help from mathematics. After several attempts, Escher found the solution in an article written a few years earlier by the geometer H. S. M. Coxeter. By viewing the early versions of *Heaven and Hell* that are on display, visitors can chart the aesthetic evolution of Escher’s tessellations.

Escher’s work is often presented as an academic exercise in visual mathematics, but his interest lay in the real world. He declared: “We are not playing a game of imaginings ... we are conscious of living in a material,



three-dimensional reality.” This is emphasized in his tree sketches, which show how the patterns of branches repeat at different scales and distort when reflected in the rippled surface of a pond.

The patterns in *Heaven and Hell* do not replicate those found in nature — their scalings shrink at a different rate. Nature produces fractal structures, as shown in one exhibit comprising a spherical mirror positioned at the centre of a cube of mirrors. Coloured light rays bounce around the mirrors, their many reflections setting up a fractal pattern when viewed from outside. But Escher did not depict this geometry, instead using the hyperbolic one described in Coxeter’s article.

Intriguingly, *Heaven and Hell* was created many years before Benoît Mandelbrot’s book *The Fractal Geometry of Nature* (W. H. Freeman, 1982), which made nature’s scaling properties well known. The exhibition thus asks to what extent Escher knew about these natural rules. The artist hints that he was conscious of them and chose alternatives: “The reality around us ... is too common, too dull, too ordinary for us. We hanker after the unnatural or supernatural, that which does not exist, a miracle.” Perhaps he achieved this miracle in what he referred to as the “deep, deep infinity” of his repeating patterns. ■

Richard Taylor is associate professor of physics, psychology and art in the Department of Physics at the University of Oregon, Eugene, Oregon 97403, USA.
e-mail: rpt@uoregon.edu

M. C. ESCHER'S CIRCLE LIMIT IV © 2009 THE M. C. ESCHER CO., THE NETHERLANDS. ALL RIGHTS RESERVED. WWW.MCESCHER.COM

Correction

The book review ‘Cooking debate goes off the boil’ by Pat Shipman (*Nature* **459**, 1059–1060; 2009) incorrectly stated that carbohydrates can be obtained from fat and marrow when, in fact, it is fat that is obtained from these sources.

NEWS & VIEWS

ECOLOGY

Production in pristine lakes

Jonathan J. Cole

An investigation of lakes in Sweden has delivered results that run counter to the idea that primary production is generally limited by the availability of nutrients. There are lessons for limnologists in this.

On page 506 of this issue, Karlsson *et al.*¹ show that, in lakes, the primary production of algae and the growth of bottom-dwelling (benthic) invertebrates and fishes is not controlled by nutrients. Rather, the determining factor is the light climate, which is itself determined by dissolved organic matter from the watershed. This conclusion is both intriguing and likely to prove controversial.

Primary production is the rate at which organic matter is synthesized from raw materials — carbon dioxide, water, inorganic nutrients — and light energy, and forms the base of the food web in most environments. In aquatic systems, excessive primary production — known as eutrophication and taking the form of algal blooms — has a host of undesirable effects, including bottom-water anoxia, loss of fish species and large-scale production of algal toxins². The main cause of eutrophication in lakes is the input of phosphorus and nitrogen, which tend to be the limiting nutrients in that demand for them by algae often exceeds supply^{3,4}. If you add enough phosphorus and nitrogen, most lakes will experience blooms. If the nutrient input is reduced, these blooms abate.

In lakes with modest to high nutrient levels, there is also a good correlation⁵ between total phosphorus (TP) and the amount of algal biomass (which is usually expressed in terms of the concentration of the photosynthetic pigment, chlorophyll *a*; this is not a direct measure of primary production but makes for a good surrogate). The TP–chlorophyll correlation is driven by the fact that phosphorus is usually the primary limiting nutrient: unless there are other complications, such as deep mixing or food-web changes, much of the available phosphorus is assimilated by algae.

All in all, there is firm evidence on multiple scales and using multiple approaches⁴ that nutrients control eutrophication. But what is the story for lakes in the pristine, low-nutrient (oligotrophic) state? When nutrients are added, not only does primary production increase, but the entire community of primary producers and consumers is also altered, often drastically. Thus, nutrient additions stimulate primary production and select for species that



JENNY ASK

Figure 1 | Clear difference. This example of a subarctic lake is nutrient-poor but productive because of a high-light climate stemming from the clear and shallow water¹.

respond to nutrients. The question here is what limits the ambient community in nutrient-poor lakes.

The somewhat surprising answer to emerge from Karlsson and colleagues' study¹ is that the light climate, and not nutrients, is the main factor controlling both primary production and secondary production of benthic invertebrates and fishes in pristine lakes (Fig. 1). Furthermore, the light climate is controlled largely by the input of highly coloured dissolved organic matter (CDOM) that enters lakes from surrounding forests. This organic matter, rich in dark-brown humic substances, gives lake water its characteristic tea-like colour, which can be very weak in some lakes but extremely dark in others.

So what limits productivity in nutrient-poor lakes, according to Karlsson *et al.*, is the input of dissolved organic matter from the surrounding watershed. Their evidence comes from a well-crafted correlation analysis of a series of

12 Swedish lakes that varied in CDOM and nutrients. Neither primary production nor the secondary production of fishes or invertebrates was positively correlated with nutrients; all were inversely correlated with the fraction of surface light that reached the mean depth of the lake, a combined effect of CDOM and depth.

Why this difference? There are strong correlations between nutrients and primary production (or rather chlorophyll *a*) and fish biomass⁶ in lakes in general. But the conclusion of this new study¹ is that the absorbance of light by CDOM is the limiting factor.

Karlsson *et al.*¹ point out that many lakes are small, nutrient-poor, rather shallow and have significant amounts of benthic primary production in addition to the phytoplankton production that occurs in the open water above the lake bottom. Indeed, globally, most lakes are like those studied here rather than the high-nutrient lakes that dominate most TP–chlorophyll data sets. Moreover, the TP–chlorophyll

relationship includes only the planktonic component, and that just in the surface water. The strength of the relationship is driven largely by eutrophic lakes — lakes with high TP, such as those already experiencing anthropogenic eutrophication. For pristine lakes there is much more variation in CDOM than there is in TP (ref. 7). Unless there is some large external input of phosphorus, CDOM will explain more of the variation in primary production than will TP.

Should we thus conclude that we need not worry about nutrient inputs to lakes? The answer is ‘no’. If large amounts of nutrients were added to oligotrophic lakes, the lakes would, in all likelihood, also experience algal blooms, as do other humic lakes^{8,9}. The study¹ does suggest that the anthropogenic and climatic factors that are affecting the concentrations of CDOM

in many parts of the world¹⁰ will probably have an effect on primary production in small lakes. But although the authors’ conclusions are both bold and reasonable, the data come from only a handful of lakes, and all of them in Sweden. The idea that light and CDOM limit primary production will remain controversial until it is tested against a larger and more diverse array of lakes.

There are, however, lessons for limnologists in this paper. Pay attention to CDOM. Consider both benthic and pelagic production of the whole lake (not just the surface water). And do not assume that the well-known principle of nutrient limitation¹¹ can be applied in general — most of the world’s lakes contain only low quantities of both nitrogen and phosphorus, and many have lots of CDOM. ■

Jonathan J. Cole is at the Cary Institute of Ecosystem Studies, Millbrook, New York 12545-0129, USA.
e-mail: colej@caryinstitute.org

1. Karlsson, J. et al. *Nature* **460**, 506–509 (2009).
2. Rabalaïs, N. N., Turner, R. E. & Wiseman, W. J. Jr *Annu. Rev. Ecol. Syst.* **33**, 235–263 (2002).
3. Likens, G. E. *Limnol. Oceanogr. Spec. Symp.* **1**, 1–328 (1972).
4. Smith, V. H. in *Successes, Limitations, and Frontiers in Ecosystem Science* (eds Pace, M. L. & Groffman, P. M.) 7–49 (Springer, 1998).
5. Smith, V. H. *Limnol. Oceanogr.* **27**, 1101–1111 (1982).
6. Hanson, J. M. & Leggett, W. C. *Can. J. Fish. Aquat. Sci.* **39**, 257–263 (1982).
7. Carpenter, S. R., Cole, J. J., Kitchell, J. F. & Pace, M. L. *Limnol. Oceanogr.* **43**, 73–80 (1998).
8. Schindler, D. W. et al. *Proc. Natl Acad. Sci. USA* **105**, 11254–11258 (2008).
9. Smith, V. H. & Schindler, D. W. *Trends Ecol. Evol.* **24**, 201–207 (2009).
10. Monteith, D. T. et al. *Nature* **450**, 537–540 (2007).
11. Sterner, R. W. *Int. Rev. Hydrobiol.* **93**, 433–445 (2008).

QUANTUM MECHANICS

Hidden context

Boris Blinov

The idea that physical phenomena might be described by a more down-to-earth theory than quantum physics has met with resistance from many physicists. Indeed, it seems that nature is not as simple as we would like.

Quantum mechanics is weird, there is no doubt about that. Efforts to demystify this theory and reconcile it with the laws of classical physics have led to the development of theories of hidden local variables — the hypothetical parameters of a quantum system that govern the seemingly random outcomes of quantum measurements. As it turns out, these theories require quantum mechanics to be non-contextual: the results of measurements on a quantum system should not depend on the act and specific properties (the context) of the measurement process itself; they should be set prior to it. But on page 494 of this issue, Kirchmair and colleagues¹ report the results of an experiment on a system of trapped ions and demonstrate that quantum mechanics conflicts with non-contextuality regardless of the quantum state of the system. This is a significant advance, coming on the heels of recent tests of quantum contextuality performed with photons² and neutrons³ in very special quantum states.

The concept of non-contextuality can be understood in simple terms. Imagine you shake a pair of dice in a cup and then flip the cup bottom-up on a table. The dice are hidden from you, but they have stopped rattling around and the values of their top faces are now set. These values are random (given that the dice are fair), but from the moment you place the cup on the table they are fixed (Fig. 1a). If you now remove the cup and see the outcome, it will be the same as the value hidden inside the cup (Fig. 1b). You could do the same exercise tomorrow, or

you could look at one die first and at the other die next, or vice versa, and the outcome would still be the same. This ‘classical’ measurement process is thus non-contextual.

But ‘quantum dice’ behave differently. Until the moment you observe them, the values of the quantum dice are not set, even though they have stopped rattling. The top face of each die is a superposition of the one-, two-, three-, four-, five- and six-dot faces (Fig. 1c). It is only when the cup is removed and the quantum dice are observed that a certain value is assigned to the top face (Fig. 1d). What’s more, the observed values depend on the way you measure their states — that is, on how you look at them: it may be different if you look at the dice one after another or at the same time. The values are revealed only in the particular context of the measurement.

The celebrated Bell’s inequality⁴, the violation of which by quantum systems has been used to refute hidden-variable theories, requires the preparation of a very specific quantum state called the entangled state. But, in general, other types of quantum state (there are many!) do not violate this inequality. A definitive test of non-contextuality would involve demonstration of the violation of a Bell-like inequality for any quantum state, not just the entangled state. Such a test entails many more sequential operations and measurements on the quantum states than Bell-inequality tests. Moreover, the measurements themselves must be compatible, which means that they must be done simultaneously or, if done sequentially, should

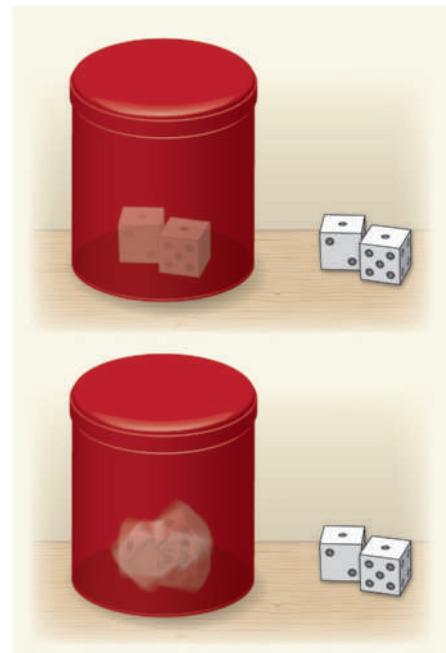


Figure 1 | Quantum dice. **a, b**, If a pair of dice is shaken in a cup, the dice stop spinning as soon as the cup is turned over on a table. The values observed on the top faces of the dice when the cup is lifted from the table are the same as the values hidden inside the cup. These ‘classical’ dice are non-contextual — that is, their values are predetermined, and are unaffected by the way we look at them, whether we observe them simultaneously or sequentially. **c, d**, By contrast, ‘quantum’ dice do not have a predetermined value. In a way, they continue spinning until they are observed. The circumstances under which they are observed — the context of the measurement process — affect the outcome of the measurement. Kirchmair and colleagues¹ use a pair of trapped ions as their quantum dice to demonstrate the contextual nature of quantum mechanics.

be unaffected by the order in which they are performed.

In their experiment, Kirchmair and colleagues¹ used a pair of trapped ions (their

quantum dice) for which almost any spin state could be prepared on demand. Following this spin-state preparation step, the authors made a sequence of compatible measurements on the spin states to test whether a Bell-like inequality was obeyed or violated. Each sequence included measurements of various spin components. Remarkably, for all quantum states tested, the authors observed a strong violation of the inequality — a clear demonstration of the contextual nature of quantum mechanics.

One special feature that allowed Kirchmair *et al.* to accomplish these sequential compatible measurements is the technique known as quantum non-demolition measurement. The first thing one usually learns about quantum states is that they can be superpositions of various ‘basis states’, just like Schrödinger’s cat, which can be dead and alive at the same time. The second thing is that a measurement of a quantum state always results in the demolition, or ‘collapse’, of the superposition: the poor cat turns out to be dead or alive, but not both, as soon as we observe it. By contrast, the quantum non-demolition measurement does not destroy the superposition — to a degree, any following compatible measurement remains a valid quantum measurement with a random outcome.

Of course, no experiment is ever perfect, and experimental limitations can leave open various loopholes. For example, the extent to which Kirchmair and colleagues’ measurements were truly compatible was limited by the precision with which the spin states were prepared, manipulated and measured. But as far as quantum-contextuality experiments are concerned, the authors have closed a crucial loophole — the detection loophole. In their experiment, the spin states of the trapped ions were detected with an efficiency exceeding 99%, much greater than the loophole demands. Another possible loophole that has to do with imperfections in the control and manipulation of the ion spins was studied in detail by the authors for one particular spin state. They found that these imperfections did not affect the legitimacy of their findings: the non-contextual nature of the quantum state remained invalid. Extending this detailed study to other possible states is a challenge for the following generation of experiments probing quantum contextuality.

So what’s next for quantum contextuality? I believe we will soon witness the results of more precise tests of contextuality in different quantum systems. Performing contextuality tests in quantum systems whose components are separated by macroscopic distances, such as the system of two remotely located trapped ions⁵, would help to close the ‘locality loophole’. Studies that involve closing all major loopholes in a single experiment⁶ also offer a promising route. Applications of the contextual nature of quantum mechanics may include new forms of quantum cryptography — for example, one that involves a three-state quantum system as the

carrier of information⁷ — or other quantum-information-processing technologies. ■
Boris Blinov is in the Department of Physics, University of Washington, Seattle, Washington 98195, USA.
e-mail: blinov@uw.edu

1. Kirchmair, G. *et al.* *Nature* **460**, 494–497 (2009).

2. Huang, Y.-F., Li, C.-F., Zhang, Y.-S., Pan, J.-W. & Guo, G.-C. *Phys. Rev. Lett.* **90**, 250401 (2003).
3. Bartosik, H. *et al.* Preprint at <http://arxiv.org/abs/0904.4576> (2009).
4. Bell, J. S. *Physics* **1**, 195–200 (1964).
5. Moehring, D. L. *et al.* *Nature* **449**, 68–71 (2007).
6. Rosenfeld, W. *et al.* Preprint at <http://arxiv.org/abs/0906.0703> (2009).
7. Bechmann-Pasquinucci, H. & Peres, A. *Phys. Rev. Lett.* **85**, 3313–3316 (2000).

MATERIALS SCIENCE

Soft particles feel the squeeze

Daan Frenkel

It’s hard to fit in when you’re different — especially if you’re a large particle trying to squeeze into an array of smaller ones. But some soft, polymeric particles simply shrink to fit the space available.

A crystal can grow only if all of its atomic or molecular building blocks fit into the periodic lattice. This is true even for colloidal crystals, which form through the ordered self-assembly of micrometre-sized particles. The requirement for periodicity puts stringent constraints on the variation in the size of particles that can be incorporated into a given colloidal crystalline lattice. But reporting in *Angewandte Chemie*, Iyer and Lyon¹ show that crystals made of microgel particles are much more tolerant of particle size variations than was expected. This surprising feature might have practical implications for the design of ordered colloidal materials.

Crystalline materials are so familiar that we rarely pause to think about how remarkable they are. Every crystal of table salt consists of as many as 10^{18} sodium and chloride ions, all of which self-assemble into a nearly perfect periodic lattice. Yet atoms or molecules can form crystals only if their size distribution is sufficiently narrow. For crystals made up of one kind of atom (such as solid argon), or even for simple salts such as sodium chloride, the restrictions are irrelevant, because all particles that occupy equivalent lattice sites have exactly the same size. But as the metallurgist William Hume-Rothery pointed out in the 1930s, there are strong constraints on the relative atomic sizes of metals that, when mixed together, can form substitutionally disordered crystals (in which equivalent lattice sites aren’t necessarily occupied by the same kinds of atom). If the radius of two otherwise similar metal ions differs by more than 15%, the metals will not mix at all in the crystalline phase, even though they may do so in a melt.

For colloidal crystals, something similar holds. Hard, uncharged colloidal particles can form substitutionally disordered crystals of a particular cubic arrangement — face-centred cubic, to use the jargon — only if the natural spread in the sizes of the particles is less than 5.7%. If the particle size variation of a colloid is

larger than this critical value, the particles can still form crystals, but each crystal will contain particles within a certain size range, such that the variation within every crystal is again less than 5.7%. Substitutionally ordered colloidal crystals (in which equivalent lattice sites contain the same kind of particle) may also form from two different kinds of particle that have very specific size ratios; but in these cases, there

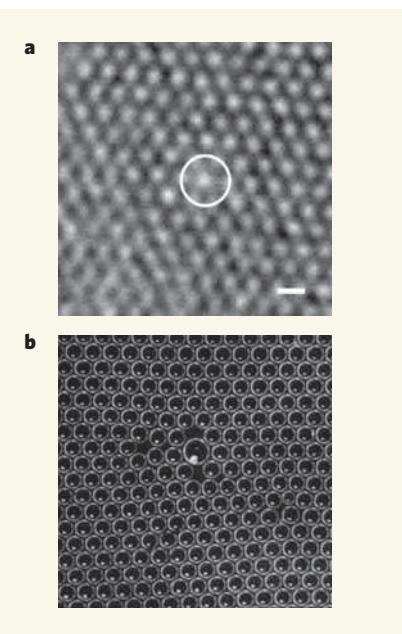


Figure 1 | Squeezing in. **a**, Iyer and Lyon¹ report that arrays of soft microgel particles, such as the array shown in this micrograph, tolerate the presence of foreign microgel particles (circled) that are much larger than their hosts. The foreign particles shrink to fit the space available, without disrupting the lattice. Scale bar, 1 μm. (Image taken from ref. 1.) **b**, By contrast, the introduction of a large, deformable particle into an array of smaller deformable particles usually causes substantial local distortion of the array, as shown in this two-dimensional raft of bubbles. The diameter of the uniform bubbles is about 1.3 mm. (Image taken from ref. 2.)

are constraints on the size variations of each constituent particle type.

Given that most crystals are rather intolerant of the inclusion of particles of significantly different sizes, Iyer and Lyon's observations¹ are a surprise. First, the authors find that colloidal crystals of soft 'microgel' particles can incorporate foreign microgel particles of a very different size from those in the host lattice — the volume of the foreign particles may be up to 15 times that of the host particles. But equally surprising is that the oversized particles fit into the host structure without causing any observable deformation of the surrounding crystal lattice (Fig. 1a). Instead, the large particles shrink until their size is comparable to that of the host particles, in what could be described as colloidal mimicry.

This mimicking behaviour is highly unusual, even for crystals made up of deformable building blocks. An early example² of such a system was studied in 1947 by Lawrence Bragg and John Nye, who investigated the properties of 'bubble rafts'. They observed that, if the sizes of all the bubbles in a two-dimensional raft are equal, the bubbles assemble in a highly ordered triangular lattice (Fig. 1b). If a much larger bubble is introduced into the raft, it causes substantial local distortion of the crystal lattice, even though the rogue bubble could in principle have deformed to fit the space available.

Colloidal crystals of ultra-soft microgels behave differently from bubble rafts because microgel particles are not only extremely deformable, they can also swell or shrink dramatically, depending on the external conditions (such as temperature, ionic strength or pH). The physical origin of this unusual trait is related to the molecular structure of microgel particles, which consist of weakly crosslinked polymer networks. A delicate balance exists between the osmotic pressures in each particle — pressures caused by polymer mixing, by the presence of confined ions in the particles and by elastic deformation of the polymer network. The outcome of the interplay between these pressures determines whether the particles swell or collapse.

Forces also exist between the microgel particles. These forces tend to compress the particles, but Iyer and Lyon¹ argue that such direct contact effects are less important than the osmotic effects that are transmitted through the solvent. One could say that the particles adjust their volume because of 'osmotic peer pressure' rather than because of direct interactions between the particles.

Nevertheless, the mechanism that causes microgel particles to shrink dramatically as they are packed more densely is, to the best of my knowledge, only partially understood. Theoretical studies have been made of the osmotic collapse of individual microgel particles³ and of the effective interaction between microgel particles of a fixed radius⁴. More recently, it was predicted⁵ that ensembles of microgel particles should undergo a succession

of freezing and melting transitions as the density of the system increases. These predictions are based on models that did not take particle shrinkage into account; it is tempting to speculate that, if density-dependent shrinkage were included in the model, even more unusual phase behaviour would result. For example, under certain circumstances, a liquid-crystal-like phase (known as a hexatic phase) might be possible^{6,7}.

Iyer and Lyon's work¹ is of great interest to materials scientists, because it suggests that microgel crystals tolerate the presence of small quantities of other microgels. This would make it easier to make microgel-based photonic

materials, substitutionally doped with particles that have very different optical properties from the host microgel. ■

Daan Frenkel is in the Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, UK. e-mail: df246@cam.ac.uk

1. Iyer, A. S. & Lyon, L. A. *Angew. Chem. Int. Edn* **48**, 4562–4566 (2009).
2. Bragg, L. & Nye, J. F. *Proc. R. Soc. Lond. A* **190**, 474–481 (1947).
3. Levin, Y. *et al.* *Phys. Rev. E* **65**, 036143 (2002).
4. Denton, A. R. *Phys. Rev. E* **67**, 011804 (2003).
5. Gottwald, D. *et al.* *J. Chem. Phys.* **122**, 074903 (2005).
6. Bladon, P. & Frenkel, D. *Phys. Rev. Lett.* **74**, 2519–2522 (1995).
7. Chou, T. & Nelson, D. R. *Phys. Rev. E* **53**, 2560–2570 (1996).

CANCER

Three birds with one stone

Franck Toledo and Boris Bardot

The core domain of the p53 protein has been found to affect microRNA processing — its third known antitumour activity. Most cancerous p53 mutations affect this domain and may abolish all tumour-suppressor functions.

The p53 protein is a major tumour suppressor. The gene that encodes it (*TP53*) is mutated in about half of human cancers, and most remaining tumours find other ways to inactivate the p53 pathway¹. p53 functions mainly as a transcription factor, and responds to stresses such as DNA damage by binding to specific DNA sequences and activating target genes that promote programmed cell death (apoptosis) or permanent cell-cycle arrest (senescence). Most *TP53* mutations in cancers affect the protein's core DNA-binding domain and abrogate p53's transcriptional activity¹. Several regulators of apoptosis also interact with p53 through the DNA-binding domain, leading to the proposal² that mutations in this domain act as 'dual hits', affecting both p53's transcriptional capacity and its ability to promote apoptosis in a transcription-independent manner. Suzuki *et al.*³ now identify a third function for p53's DNA-binding domain — modulation of the processing of small RNAs called microRNAs (miRNAs).

Small noncoding miRNAs (20–25 nucleotides) are negative regulators of mammalian gene expression, and evidence that they have a role in cancer is accumulating⁴. miRNAs bind to protein-coding messenger RNAs in a sequence-specific manner, reducing mRNA stability, inhibiting translation, and thereby decreasing expression of the target protein. Mature miRNAs are formed by two sequential processing reactions — primary transcripts of miRNA genes (pri-miRNAs) are first cleaved into hairpin-shaped intermediates (pre-miRNAs) by the Drosha microprocessor complex, and pre-miRNAs are then processed into mature miRNAs by the protein Dicer.

Human cancer cells are often associated with a widespread decrease in miRNAs, and the authors reasoned that this reduction might imply that there is a direct link between tumour-suppressor pathways and the protein machinery that processes these RNAs. There is already a known association between p53 and miRNAs — p53 activates transcription of the miR-34 gene family, leading to production of miRNAs that promote apoptosis or senescence⁵. Suzuki *et al.*³ find that, after DNA damage, p53 interacts with the Drosha complex, facilitating the processing of a subset of pri-miRNAs to pre-miRNAs. The authors identify six members in this new category of p53-regulated miRNAs, and show that, in cancer cells, overexpression of each of these miRNAs decreases the rate of cell proliferation. Although these data rely on miRNA over-expression, it is reasonable to assume that, in a more physiological setting, a modest but concerted upregulation of several miRNAs by p53 might contribute to its antitumour response.

Suzuki *et al.*³ show that the DNA-binding domain of p53 is required for its interaction with Drosha, and that p53 probably binds to the p68 RNA-helicase component of the Drosha complex. The authors find that three types of missense mutation in the DNA-binding domain of p53 that obliterate its transcriptional activity also lead to decreased processing of pri-miRNAs by Drosha. They therefore propose that p53 mutants might reduce the interaction between pri-miRNAs and Drosha-complex proteins. Moreover, they provide data suggesting that mutant p53 interferes with the association between p68 and the Drosha complex, perhaps by titrating p68 away from Drosha.

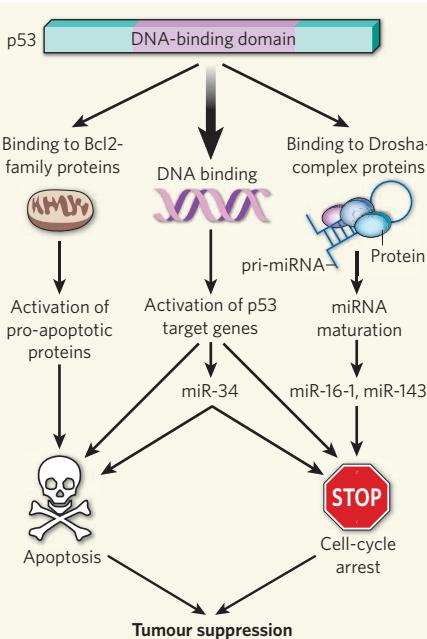


Figure 1 | Three antitumour functions. The DNA-binding domain of p53 lies in the core of the protein and has three antitumour functions. It binds to DNA and enables the activation of target genes (including the miR-34 gene family) to induce apoptosis or cell-cycle arrest; it stimulates apoptosis through an interaction with proteins of the Bcl2 family at the mitochondrion; and, as Suzuki *et al.*³ show, it interacts with proteins of the Drosha complex to promote processing of a subset of miRNAs, including miR-16-1 and miR-143, which suppress cell proliferation. Most p53 mutations in human cancers lie in the DNA-binding domain and may affect all three functions.

Interestingly, p53 mutants that are frequently observed in human tumours — such as R175H and R273H — correlate with lower Drosha activity, whereas mutant C135Y, which is rarely found in cancers, has apparently less effect on Drosha³. Studies of additional p53 mutants will be required to further test whether there is an inverse correlation between the frequency of occurrence of a specific mutant p53 in human tumours and the magnitude of its effect on Drosha activity.

Mice with mutant p53 (ref. 6) have provided compelling evidence that some p53 mutants do not simply lose normal tumour-suppressor functions, but also acquire cancer-promoting (oncogenic) properties. Suzuki and colleagues' suggestion³ that p53 mutants might reduce the interaction between pri-miRNAs and Drosha-complex proteins provides a possible, and previously unsuspected, mechanism by which mutant p53 could induce cancer. Together with previous studies^{1,2}, these findings suggest that, in human cancers, mutations that affect the DNA-binding domain of p53 essentially perform a hat-trick by hitting three tumour-suppressive functions at once: activation of target genes, induction of transcription-independent apoptosis, and processing of a subset of miRNAs (Fig. 1).

It will be important to establish the full repertoire of miRNAs that are upregulated by p53 through the mechanism described by Suzuki *et al.*, because this may provide clues about gene products whose function must be suppressed to promote tumour formation. Another member of the p53 family, p63, seems to be the main regulator of the DNA-damage response in female germ cells⁷ — which were not studied by Suzuki and colleagues. It will be interesting to determine whether p63 regulates miRNA processing in germ cells.

TP53 is transcribed into nine different mRNAs, some of which are misregulated in human cancers⁸. Little is known about these different transcripts, including whether they are all efficiently translated into protein. However, the proteins they encode should contain most of the p53 DNA-binding domain, and so should have the potential to participate in the regulation of miRNA processing. The recent discovery of single nucleotide polymorphisms (SNPs) in several genes of the p53 pathway⁹ adds another level of complexity. For example, a SNP in the promoter of *MDM2*, which encodes a major p53 inhibitor, alters p53 levels and affects the age at which people get certain tumours and the survival of patients after anticancer therapies¹⁰. Because it alters the amounts of p53, the same SNP probably also affects miRNA processing. This suggests the existence of distinct efficiencies of miRNA processing in human populations that may account, in part, for differences in the age and frequency of cancer onset or prognosis.

As more and more functions are ascribed to p53 (ref. 11), the mechanism identified by Suzuki *et al.*³ has implications beyond the cell's response to DNA damage and cancer. The data suggesting that mutant p53 titrates the p68 RNA helicase³ are particularly intriguing, because p68 is a transcriptional co-regulator that is also involved in RNA splicing¹². Hence, mutant p53 could alter other aspects of RNA metabolism besides miRNA processing. In the ever-expanding universe of p53, the regulation of gene expression by p53–Drosha/p68 interactions may well be the next Big Bang. ■

Franck Toledo and Boris Bardot are in the Genetics of Tumor Suppression Group, Institut Curie, 75248 Paris Cedex 05, France, and at the Université Pierre et Marie Curie Paris 06, Paris, France.
e-mail: franck.toledo@curie.fr

- Toledo, F. & Wahl, G. M. *Nature Rev. Cancer* **6**, 909–923 (2006).
- Green, D. R. & Kroemer, G. *Nature* **458**, 1127–1130 (2009).
- Suzuki, H. I. *et al.* *Nature* **460**, 529–533 (2009).
- Ventura, A. & Jacks, T. *Cell* **136**, 586–591 (2009).
- He, L., He, X., Lowe, S. W. & Hannon, G. J. *Nature Rev. Cancer* **7**, 819–822 (2007).
- Lozano, G. *Curr. Opin. Genet. Dev.* **17**, 66–70 (2007).
- Suh, E.-K. *et al.* *Nature* **444**, 624–628 (2006).
- Bourdon, J.-C. *et al.* *Genes Dev.* **19**, 2122–2137 (2005).
- Whibley, C., Pharoah, P. D. P. & Hollstein, M. *Nature Rev. Cancer* **9**, 95–107 (2009).
- Vazquez, A., Bond, E. E., Levine, A. J. & Bond, G. L. *Nature Rev. Drug Discov.* **7**, 979–987 (2008).
- Vousden, K. H. & Lane, D. P. *Nature Rev. Mol. Cell Biol.* **8**, 275–283 (2007).
- Fuller-Pace, F. V. *Nucleic Acids Res.* **34**, 4206–4215 (2006).



50 YEARS AGO

Puzzle-Math. By Dr. George Gamow and Dr. Marvin Stern — Books of the 'mathematics for fun' type are often neither very mathematical nor very funny, but those who know some of Dr. Gamow's earlier writings will expect this volume, in spite of its catchpenny title, to combine amusement with instruction, and they will not be disappointed. The thirty-three problems are entertainingly set out, and solved by honest mathematical processes, involving little or no manipulative technique. There are some 'chestnuts'... but many of the problems are new or not widely known... A bright student might easily be led to a better appreciation of the fundamental logic of mathematics by reading this cheerful little book.

From *Nature* 25 July 1959.

100 YEARS AGO

In the July number of the *Reliquary*, Mr. E. H. Goddard continues the useful series of articles dealing with local collections of antiquities, his subject being Roman objects discovered in Wiltshire. Though the county possesses no Roman sites ranking in interest and importance with those of Dorchester, Silchester, Bath, or even Lydney or Woodchester, it contains Cunetio near Marlborough, villas at Box, Colerne, and Wraxall, and, in particular, Old Sarum.

...

On October 21, 1638, the Devil visited Widdecombe Church ... in Dartmoor, a full account of which remarkable event is recorded on a tablet in curious versification, the work of the village schoolmaster, which is preserved in the church. As a matter of fact, the place was the scene of a terrible thunderstorm, which caused the loss of several lives, damaged the tower, and caused such consternation that it was attributed to demoniacal agency. The original tablet, a curious instance of the popular beliefs current at the time, is reproduced by Mr. Le Blanc Smith in the July number of the *Reliquary*. From *Nature* 22 July 1909.

50 & 100 YEARS AGO

GEOMORPHOLOGY

Landscape texture set to scale

Kelin X. Whipple

Why, in many landscapes, does ridge-valley spacing show such regularity? The combination of high-resolution data and an elegant model offers a solution to this long-standing puzzle, for some cases at least.

The dissection of topography into ridges and valleys sets the scene for all manner of processes — physical, chemical and biological — that operate on Earth's surface. How do these processes interact to set the scale of this dissection? Although many other aspects of landforms seem to be scale-invariant¹, there are plenty of instances in which our own eyes tell us that ridges and valleys are uniformly spaced (Fig. 1) — that there is a fundamental scale to landforms. This impression can be confirmed quantitatively².

For more than a century, geomorphologists have recognized that valley spacing is probably governed by a transition from hillslope (unchannelled) soil transport to channelled stream incision. Hillslope soil transport acts in a diffusive manner and tends to fill in incipient channels cut by stream flow, and competition between these two processes has long been thought to determine the scale of landscape dissection. Much progress has been made^{3–6},

but no general theory with demonstrably predictive power has emerged — until now, that is, for Perron *et al.*⁷ (page 502 of this issue) show that the wide variation in ridge-valley spacing can be explained by a simple model of this competition in a certain class of landscape.

From a scaling analysis of a statement of the conservation of mass, in previous work Perron *et al.*⁶ derived a non-dimensional quantity, akin to a value called the Pécelt number, that gauges the competition between soil creep and channel incision. Their numerical simulations demonstrated that this non-dimensional quantity does indeed, in principle, govern valley spacing.

In their latest paper, Perron *et al.*⁷ test this model against observations using newly available, high-resolution topographic data⁸ that reveal a tenfold variation in valley spacing among five field sites. To define a simple, testable hypothesis, they restrict their analysis to 'low-relief, soil-mantled' landscapes, where

soil creep and stream flow are the dominant processes controlling sediment transport and erosion. For these cases, their model and their scaling analysis predict that valley spacing scales with the ratio of the intensity of soil creep to the intensity of channel incision (D/K in their notation), but interestingly not with the actual erosion rate — a prediction in close agreement with previous theoretical work^{4,5}. Perron *et al.* have developed a new method that allows the D/K ratio to be estimated directly from topographic data. They find a more than 60-fold variation in D/K among the field sites in different climatic and geological settings that correlates strongly with independently measured valley spacing in a manner consistent with model predictions.

But what sets the ratio of intensities of soil creep and channel incision (D/K)? Both the denominator and the numerator in this ratio are known, qualitatively, to depend in various ways on substrate properties, climate, hydrology and biota. However, the linkages among these factors as they co-evolve are notoriously complex. As yet no theory exists to predict how D/K should vary with climate or rock type, and new data are needed to guide fresh thinking.

Nonetheless, despite some scatter, the data used in this study⁷ do provide some clues. They suggest that weaker rocks and drier climates are associated with closely spaced valleys (channel incision dominant), whereas stronger rocks and wetter climates are associated with widely spaced valleys (soil creep dominant) (Fig. 2). These findings are striking, in two respects.

First, although it is the less robust indication, the hint of a climatic control is interesting because the direct dependence of the intensity of channel incision on run-off^{4,5} intuitively implies the opposite trend — that wetter climates would be associated with lower D/K and more closely spaced valleys. However, greater vegetative cover and increased soil disturbance by vegetation or animals can reduce run-off while accelerating soil creep⁹.

Second, the tenfold difference in D/K ratio, and associated fivefold difference in valley spacing, between two of the field sites — Dragon's Back and Gabilan Mesa (Fig. 1) — is surprising. The sites are separated by only about 100 kilometres, experience a similar climate and are cut into the same moderately consolidated sedimentary rocks. These rocks seem to be slightly weaker at Dragon's Back, consistent with the general trend between rock strength and valley spacing. Dragon's Back is also slightly drier, and this climatic difference



Figure 1 | Uniformity in ridge-valley spacing. This is Gabilan Mesa, California, one of the five 'low-relief, soil-mantled' sites that provided data for Perron and colleagues' analysis⁷. The valleys are cut into moderately consolidated sediments, with the distance between them being about 160 metres.

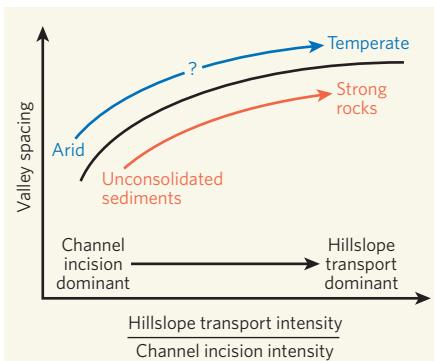


Figure 2 | Tentative interpretation of geological and climatic influences on valley spacing. According to theory, valley spacing in low-relief, soil-mantled landscapes is set by the ratio between hillslope transport (soil creep) intensity and channel incision intensity^{4–7}. Perron and colleagues⁷ provide a successful test of theory, and their data contain further hints about the underlying controls exercised by the substrate and climate: weaker rocks and (less certainly) drier climates seem to lead to a dominance of channel-incision processes and more closely spaced valleys.

is associated with a change from semi-arid grassland to oak savannah. Could these small differences be sufficient to explain the five-fold difference in valley spacing, or are other factors at play?

Much remains unknown about the complex controls on the intensity of soil creep and channel incision. But the work by Perron and colleagues⁷ will encourage further investigation because their ratio seems to set the fundamental length scale in landscapes, and also suggests a method for quantifying rate constants for both creep and incision. The analysis has limitations, as the authors acknowledge. It is restricted to low-relief, soil-mantled landscapes, where landslides, earth flows and debris flows do not occur; it does not account for a threshold for channel incision^{4,5}; and it is applicable only to quasi-steady-state conditions, in which erosion rate is approximately constant in space and time. Other considerations that merit attention are how temporal variations in rock uplift or climate change, over millennial to 100,000-year scales, may influence either topographic estimates of D/K or the relationship between D/K and valley spacing.

Despite these limitations, Perron *et al.*⁷ have provided a powerful tool for further exploration. It promises to help deliver additional discoveries about the complex interactions between the physical, chemical and biological processes acting on the land surface.

Kelin X. Whipple is in the School of Earth and Space Exploration, Arizona State University, Tempe, Arizona 85287-1404, USA.
e-mail: kxw@asu.edu

- Rodríguez-Iturbe, I. & Rinaldo, A. *Fractal River Basins: Chance and Self-Organization* (Cambridge Univ. Press, 2001).
- Perron, J. T., Kirchner, J. W. & Dietrich, W. E.

J. Geophys. Res. doi:10.1029/2007JF000866 (2008).

- Smith, T. R. & Bretherton, F. P. *Wat. Resour. Res.* **8**, 1506–1529 (1972).
- Howard, A. D. *Earth Surf. Process. Landf.* **22**, 211–227 (1997).
- Tucker, G. E. & Bras, R. L. *Wat. Resour. Res.* **34**, 2751–2764 (1998).
- Perron, J. T., Dietrich, W. E. & Kirchner, J. W.
- J. Geophys. Res.* doi:10.1029/2007JF000977 (2008).
- Perron, J. T., Kirchner, J. W. & Dietrich, W. E. *Nature* **460**, 502–505 (2009).
- National Center for Airborne Laser Mapping. www.ncalm.org.
- Fernandes, N. F. & Dietrich, W. E. *Wat. Resour. Res.* **33**, 1307–1318 (1997).

BIOENGINEERING

Cellular control in two clicks

Jason A. Burdick

If complex tissues are to be engineered, synthetic materials will be needed that provide cells with precisely located molecular cues. A method that attaches such cues to specific areas of a gel could be the answer.

Contrary to popular belief, George Washington's dentures weren't made of wood — they were actually made of ivory and gold. As with many other materials that have been used in biological systems, ivory and gold are relatively inert, and simply provided mechanical support for their intended application. Although giving biomaterials only a supporting role restricts their applications, it has nevertheless led to the development of a range of clinically useful materials (such as bone cements) and implantable devices (such as fixation plates for holding fractured bones in place). But what is really needed are biologically active materials that interact with and signal to surrounding cells and tissues.

Of particular interest are materials that can both track and manipulate the local three-dimensional arrangement of cells. Such an achievement would have been unimaginable

just a few years ago, yet it is exactly what has been reported by DeForest *et al.*¹ in *Nature Materials*. Using cell-compatible reactions², the authors first encapsulate cells in hydrogels — water-swollen polymer networks — and then introduce precisely targeted molecules into the gels to either monitor or alter the cells' behaviour. Hydrogels are especially attractive as media for cell culture because they provide a tissue-like, three-dimensional environment in which cells can flourish³.

The past couple of decades have seen an explosion of work in which engineers and biologists have collaborated to find ways of assembling cells, molecules and scaffold materials, with the ultimate goal of growing biological tissues⁴. These efforts are leading to new therapies, but progress has been slow, especially in producing tissues that consist of several cell types or that lack the capacity to

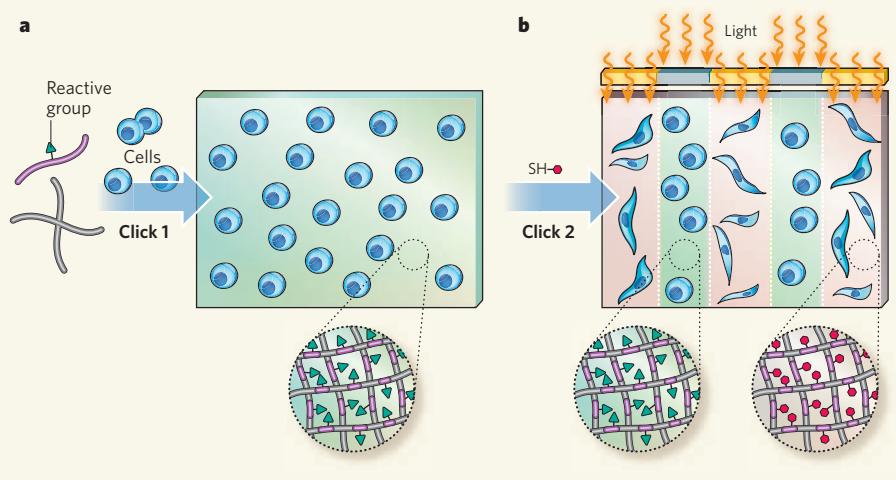


Figure 1 | Clicking into place. DeForest *et al.*¹ have developed a method for monitoring or controlling the behaviour of cells at user-defined sites within a hydrogel (a water-swollen, polymeric network). **a**, The gel is prepared from its monomers in the presence of cells using a 'click' reaction, so that the cells become encapsulated in the gel. The crosslinks of the polymer network incorporate a reactive chemical group. **b**, When irradiated with light, the chemical group reacts in a second click reaction with molecules that contain thiol (SH) groups; these molecules act as signals that monitor or dictate the behaviour of the encapsulated cells. In the example shown, light is shone through a mask, so that only groups in the illuminated regions react. Once incorporated into the gel, the signal molecules cause the cells to change shape.

regenerate after damage. The problems might stem from the lack of materials that provide appropriate biological signals for cells. *In vivo*, cells receive many cues from their surroundings⁵. Factors such as the topography of the local environment, its chemical and mechanical properties, and the availability of soluble molecules can combine to instruct cells to migrate, proliferate or — especially for stem cells — turn into another type of cell. Materials that can be manipulated in both time and space to provide such cues might be the only option if complex tissues such as heart muscle or brain tissue are to be engineered.

Some materials that can control the arrangement of cells have already been reported. Most of these use light to trigger the control processes because high spatial resolution can be obtained using laser beams or by shining light through masks. For example, materials have been made⁶ that contain chemical groups that decompose when irradiated with light, thus exposing molecules that react with cell-binding peptides. This approach enabled neural cells to grow into a hydrogel at defined locations, but has limited potential for direct cell encapsulation because the process produces potentially toxic side products. More recently, another hydrogel system was described⁷ in which distinct regions were created that either permit or inhibit the spread of cells, depending on the degree and type of crosslinking in the material. This process produces no toxic side products, but it is not compatible with many biological cues, and it also causes structural changes in the material — which can lead to inadvertent signalling to cells.

DeForest *et al.*¹ report the first biologically compatible system for encapsulating cells in hydrogels that have spatially defined cues, and in which the cues are introduced without affecting the molecular structure of the polymer network. In their approach, the authors first form a uniform hydrogel in the presence of cells, using a 'click' reaction — a specially designed reaction in which two small chemical groups combine quickly and reliably. They thus obtain a material that is highly swollen with water and has tissue-like properties (Fig. 1).

In the second step of the process, the authors use a different click reaction to incorporate biologically active molecules into the material. This reaction proceeds only where light is shone, so that molecules can be incorporated at specific regions with micrometre precision, using a laser or a mask. A wide range of molecules that act as biological cues can be incorporated in this way, as long as they contain an appropriate reactive group. For example, any peptides that contain a cysteine amino acid fall into this category, because cysteine contains a reactive thiol (SH) group.

To illustrate the general applicability of their technology, DeForest *et al.* incorporated two different molecules into their hydrogel network. The first was a peptide that fluoresces after it is cleaved by enzymes secreted by

cells. The authors thus showed that their method can be used in real time to image the enzymatic behaviour of cells in regions of the gel that can be defined by the user. The second molecule was a peptide that acts as a cellular adhesion site. It is known that, when included in synthetic networks containing adhesion sites and appropriate enzyme-degradable crosslinks, cells can spread and remodel their surroundings⁸. Using their click reaction to control the location of adhesion sites, DeForest *et al.* were able to exert spatial control of cell spreading in their gels.

This study¹ represents a giant leap towards the goal of controlling and monitoring cells in three-dimensional environments. But there is still much to be done to exploit such control for reconstructing multicellular tissues that will have functional properties. One limitation is that the synthetic schemes required are currently quite complex, which might curb the immediate widespread use of the technology.

Nevertheless, the authors' technique will be particularly useful for investigating how cells interact with their environment, and could one day form the basis of a method for manipulating cells for regenerative medicine. ■

Jason A. Burdick is in the Department of Bioengineering, School of Engineering and Applied Science, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA.
e-mail: burdick2@seas.upenn.edu

1. DeForest, C. A., Polizzotti, B. D. & Anseth, K. S. *Nature Mater.* doi:10.1038/nmat2473 (2009).
2. Baskin, J. M. *et al.* *Proc. Natl. Acad. Sci. USA* **104**, 16793–16797 (2007).
3. Cushing, M. C. & Anseth, K. S. *Science* **316**, 1133–1134 (2007).
4. Khademhosseini, A., Vacanti, J. P. & Langer, R. *Sci. Am.* **300** (5), 64–71 (2009).
5. Burdick, J. A. & Vunjak-Novakovic, G. *Tissue Eng. A* **15**, 205–219 (2009).
6. Luo, Y. & Shiochet, M. S. *Nature Mater.* **3**, 249–253 (2004).
7. Khetan, S., Katz, J. S. & Burdick, J. A. *Soft Matter* **5**, 1601–1606 (2009).
8. Lutolf, M. P. & Hubbell, J. A. *Nature Biotechnol.* **23**, 47–55 (2005).

INFECTIOUS DISEASES

An ill wind for wild chimps?

Robin A. Weiss and Jonathan L. Heeney

Simian immunodeficiency virus is associated with increased mortality in a subspecies of chimpanzee living under natural conditions in East Africa. This is worrying news for the chimpanzee populations involved.

Today's pandemic strain of HIV-1 crossed the species barrier from chimpanzees (*Pan troglodytes*) to humans less than 100 years ago. Until now, it has been widely assumed that the precursor of HIV-1, chimpanzee simian immunodeficiency virus (SIVcpz), causes little, if any, illness in its animal host. On page 515 of this issue, however, Keele and colleagues¹ show that small groups of wild chimpanzees naturally infected with SIVcpz do develop hallmarks of AIDS. Careful monitoring for almost a decade has revealed that SIV-infected animals of the eastern subspecies of chimpanzee (*Pan troglodytes schweinfurthii*) in the Gombe National Park in Tanzania have a markedly higher death rate than non-infected animals.

Chimpanzees — humans' closest living relative — are not only genetically similar to humans but also share susceptibility to some infectious diseases. For instance, outbreaks of haemorrhagic fever caused by relatives of Ebola virus in chimpanzees and gorillas have resulted in marked mortality in wild populations². However, subtle chronic diseases like those caused by lentiviruses such as SIV/HIV are more difficult to document in wild non-human primates than are acute diseases with high mortality rates.

More than 40 strains of SIV are known to cause natural infection in African non-human primates³, but few cases of AIDS have



M. SERES, AAP

Figure 1 | Survivor. Noah, the first *Pan troglodytes schweinfurthii* to be identified with naturally acquired SIVcpz infection, is alive and well almost 20 years later. By contrast, a number of SIVcpz-infected chimpanzees in the wild seem to develop an AIDS-like illness.

been recorded, and progression to illness after viral infection is thought to be rare. This lack of disease has been well documented in African green monkeys and sooty mangabeys⁴. In addition, a captive *P. t. schweinfurthii* is in good

health almost 20 years after being identified as having naturally acquired SIVcpz infection (Fig. 1). However, exceptions have been noted in primates in captivity. For example, a captive sooty mangabey developed classic symptoms of AIDS 18 years after natural infection with SIV⁵. This long incubation period exceeded the average lifespan of sooty mangabeys in the wild and, because the adaptation of SIV to its natural simian hosts would not necessarily include complete avirulence, it is difficult to draw conclusions from this isolated example. AIDS has also been documented in black mangabeys in captivity following transmission of SIVsm from sooty mangabeys, their naturally infected cousins⁴.

The outcome of infection with HIV or SIV is determined both by variants of the viral population and by individual host differences. In studies conducted largely in the 1980s, chimpanzees (mostly of the subspecies *Pan troglodytes verus*) that were infected with a particular laboratory strain of HIV-1 controlled the amount of virus in the blood, and remained healthy, although a human accidentally exposed to the same strain of HIV-1 developed AIDS⁶. A report of one case indicated that certain variants of HIV-1 can cause disease in chimpanzees⁷. However, such experiments are no longer permitted owing to the endangered status of chimpanzees.

Keele *et al.*¹ are the first to examine the long-term effects of naturally transmitted SIVcpz in wild chimpanzee populations in their natural habitat. The research was made possible because the Gombe chimpanzees have been studied in their natural setting for many decades by Jane Goodall and colleagues, and these animals are accustomed to the presence of humans. Samples of faeces collected from the forest floor can be used to detect SIVcpz DNA and the infected individual can be identified by performing genetic fingerprinting on the same sample. In addition, antibodies directed against SIVcpz can be detected in urine. Using such techniques, 94 members of social groups of chimpanzees in Gombe were meticulously followed over a 9-year period without the need to capture them or take blood samples to determine rates of SIVcpz infection and its effects. Overall, 17 (18%) apes were infected with SIVcpz. Several infected animals that died were autopsied, revealing pathology compatible with human AIDS. The SIV-infected apes had a 10- to 16-fold higher mortality rate than did non-infected apes, and a smaller proportion of infected females gave birth to offspring, none of which survived for more than one year.

SIVcpz in chimpanzees is apparently less virulent than HIV-1 is in humans. Why? One hypothesis⁸ is that SIVcpz and its chimpanzee host have co-evolved over millions of years. But SIVcpz may in fact be a relatively recent introduction into chimpanzees — it seems to be a recombinant virus with genome components from SIVs of other monkeys, such as red-capped mangabeys and guenons⁹. A recent

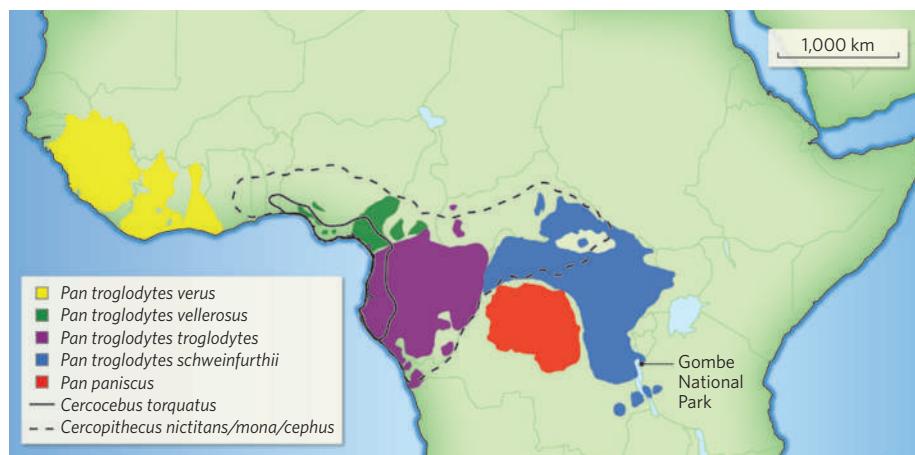


Figure 2 | Geographical distribution of the chimpanzee. There are four extant subspecies of common chimpanzee: *Pan troglodytes verus*, *Pan troglodytes vellerosus*, *Pan troglodytes troglodytes* and *Pan troglodytes schweinfurthii*. The distribution of the pygmy chimpanzee, or bonobo, *Pan paniscus*, is also shown. Only the subspecies *P. t. troglodytes* and *P. t. schweinfurthii* are known to harbour SIVcpz. Gombe National Park in Tanzania (the site of Keele and colleagues' study¹) is at the most eastern border of *P. t. schweinfurthii* territory. The precursor of SIVcpz is believed to be a genetic recombinant of lentiviruses found in two species of monkey, red-capped mangabeys (*Cercocebus torquatus*) and guenons (genus *Cercopithecus*) in west-central Africa. The west-central subspecies *P. t. troglodytes*, which harbours the SIVcpz variants that gave rise to HIV-1 of humans and SIVgor of gorillas, may have spread eastward to *P. t. schweinfurthii* causing disease.

report¹⁰ estimates that the time of the most recent common ancestor of the SIVcpz strains of all chimpanzee subspecies is 1492, although this calculation is likely to be hotly debated. Alternative explanations for differences in virulence between HIV-1 and SIVcpz could be the immune status or evolved immune genetics of the affected populations, including differences among species in host factors that restrict certain types of viral infection⁹.

The Gombe chimpanzees belong to the eastern subspecies *P. t. schweinfurthii*. The western subspecies *Pan troglodytes troglodytes*, which harbours the SIVcpz variants that have given rise to HIV-1, has not yet been identified with AIDS-like illness (Fig. 2). Could the increased virulence of SIVcpz in *P. t. schweinfurthii* be due to a recent cross-subspecies transmission of SIVcpz from *P. t. troglodytes* to an immunologically naive *P. t. schweinfurthii* population? It is notable that SIVcpz of the *P. t. troglodytes* subspecies has transmitted to humans on at least three occasions, whereas SIVcpz cross-infection to humans from *P. t. schweinfurthii* has not been documented.

In view of Keele and colleagues' results¹, why was the progression of SIV infection to AIDS-like illness not more apparent in chimpanzees in captivity? Much of the pathology of AIDS is linked to a general hyperactivation of the immune system that occurs before CD4 T-cell depletion, which is characteristic of AIDS¹¹. Certainly the parasitic, bacterial and viral burden in *P. t. schweinfurthii* chimpanzees in their natural habitat might provide the activating environment to trigger progression to AIDS in a relatively immunologically naive subspecies. General theories of the evolution of virulence, and of the relationship between virulence and pathogen dispersal¹², can help in elucidating

why certain infections cause disease whereas others don't — the abrupt changes in virulence on cross-host infection are ripe for modelling, taking into account the burgeoning knowledge of host genetic variation. Primate lentiviruses would be a promising starting point for this exercise.

The study by Keele *et al.*¹ shows the benefit of multidisciplinary research, in which primatologists, pathologists, geneticists and molecular virologists joined forces. Their work indicates that SIVcpz causes AIDS-like disease in a subspecies of chimpanzee that is already endangered in the wild. In addition to the threat of AIDS and Ebola, the great apes may also be at risk of acquiring other infections that affect humans from their increasingly close contact with our species. Relatively minor human infections could prove to be serious pathogens in chimpanzees. We are not alone. ■

Robin A. Weiss is in the Division of Infection and Immunity, University College London, London W1T 4JF, UK. Jonathan L. Heeney is in the Department of Veterinary Medicine, University of Cambridge, Cambridge CB3 0ES, UK.

e-mails: r.weiss@ucl.ac.uk; jlh66@cam.ac.uk

1. Keele, B. F. *et al.* *Nature* **460**, 515–519 (2009).
2. Bermejo, M. *et al.* *Science* **314**, 1564 (2006).
3. Aghokeng, A. F. *et al.* *Infect. Genet. Evol.* doi:10.1016/j.meegid.2009.04.014 (2009).
4. Pandrea, I., Silvestri, G. & Apetrei, C. *Curr. HIV Res.* **7**, 57–72 (2009).
5. Ling, B. *et al.* *J. Virol.* **78**, 8902–8908 (2004).
6. Beaumont, T. *et al.* *J. Virol.* **75**, 2246–2252 (2001).
7. Noverembre, F. J. *et al.* *J. Virol.* **71**, 4086–4091 (1997).
8. de Groot, N. G. *et al.* *Proc. Natl Acad. Sci. USA* **99**, 11748–11753 (2002).
9. Heeney, J. L., Dalgleish, A. G. & Weiss, R. A. *Science* **313**, 462–466 (2006).
10. Wertheim, J. O. & Worobey, M. *PLoS Comput. Biol.* **5**, e1000377 (2009).
11. Gougeon, M. L. *et al.* *J. Immunol.* **158**, 2964–2976 (1997).
12. Wild, G., Gardner, A. & West, S. A. *Nature* **459**, 983–986 (2009).

ARTICLES

Distinctive chromatin in human sperm packages genes for embryo development

Saher Sue Hammoud^{1,2}, David A. Nix³, Haiying Zhang¹, Jahnvi Purwar¹, Douglas T. Carrell² & Bradley R. Cairns¹

Because nucleosomes are widely replaced by protamine in mature human sperm, the epigenetic contributions of sperm chromatin to embryo development have been considered highly limited. Here we show that the retained nucleosomes are significantly enriched at loci of developmental importance, including imprinted gene clusters, microRNA clusters, HOX gene clusters, and the promoters of stand-alone developmental transcription and signalling factors. Notably, histone modifications localize to particular developmental loci. Dimethylated lysine 4 on histone H3 (H3K4me2) is enriched at certain developmental promoters, whereas large blocks of H3K4me3 localize to a subset of developmental promoters, regions in HOX clusters, certain noncoding RNAs, and generally to paternally expressed imprinted loci, but not paternally repressed loci. Notably, trimethylated H3K27 (H3K27me3) is significantly enriched at developmental promoters that are repressed in early embryos, including many bivalent (H3K4me3/H3K27me3) promoters in embryonic stem cells. Furthermore, developmental promoters are generally DNA hypomethylated in sperm, but acquire methylation during differentiation. Taken together, epigenetic marking in sperm is extensive, and correlated with developmental regulators.

During spermiogenesis canonical histones are largely exchanged for protamines^{1,2}, small basic proteins that form tightly packed DNA structures important for normal sperm function³. We find about 4% of the haploid genome retained in nucleosomes (Supplementary Fig. 1a). The rare retained nucleosomes in sperm consist of either canonical or histone variant proteins, including a testes-specific histone H2B (TH2B) with an unknown specialized function^{4,5}. Their presence may simply be due to inefficient protamine replacement, leading to a low random distribution genome-wide with no impact in the embryo. Alternatively, these retained nucleosomes, along with attendant modifications, might be enriched at particular genes/loci. This latter possibility would raise the possibility for programmatic retention for an epigenetic function in the embryo. To address these questions, we localized the nucleosomes retained in mature sperm from fertile donors using high-resolution genomic approaches.

Developmental loci bear nucleosomes

To address donor variability, we examined nucleosome retention in a single donor (D1) and/or a pool of four donors (donor pool). Sperm chromatin was separated into protamine-bound and histone-bound fractions. In brief, mononucleosomes were isolated (>95% yield) by sequential MNase digestion and sedimentation (Supplementary Fig. 1b–e). This mononucleosome pool was used for chromatin immunoprecipitation (ChIP; to select modified nucleosomes), or the DNA was isolated from the mononucleosome pool to represent all nucleosomes. Purified DNA was subjected to high-throughput sequencing (Illumina GAI), or alternatively, was labelled and hybridized to a high-density promoter-tiling array (9 kilobase (kb) tiled; Supplementary Fig. 2, schematic).

Our initial array approach examined three replicas of D1 (pairwise average $R^2 = 0.85$). Notably, Gene Ontology analysis revealed nucleosomes significantly enriched at promoters that guide embryonic development—primarily developmental transcription factors and signalling molecules (Gene Ontology term false discovery rate

(FDR) < 0.01; Box 1 and Supplementary Table 1; for all extended Gene Ontology categories see Supplementary Tables and Supplementary Data Set 1). To conduct genome-wide profiling, we performed high-throughput sequencing of nucleosomes from D1 or the donor pool. Regions significantly enriched for histone relative to the input control (sheared total sperm DNA) were identified using a 300-base-pair (bp) window metric⁶. For display, we depict the normalized difference score and FDR window scores (Fig. 1a, FDR transformation ($-10 \log_{10} (q\text{-value FDR})$), 20 = 0.01, 25 = 0.003, 30 = 0.001, and 40 = 0.0001). Histone-enriched loci for one individual (D1) were well correlated with a donor pool ($r = 0.7$). Globally, 76% of the top 9,841 histone-enriched regions (FDR 40 cutoff) intersect genic regions, whereas the expected intersection given random distribution is 36% ($P < 0.001$).

Interestingly, sequencing of D1 or the donor pool revealed significant ($FDR < 0.001$) histone retention at many loci important for embryo development, including embryonic transcription factors and signalling pathway components (Box 1, Supplementary Tables 2 and 3). We show this enrichment at HOX loci (Fig. 1, Supplementary Fig. 3), but also observe this at stand-alone developmental transcription factors (Supplementary Fig. 4) and signalling factors (Supplementary Fig. 5). An FDR of 60 yields 4,556 genes, of which 1,683 are grouped with developmental Gene Ontology categories (2,848 total developmental genes). The magnitude of nucleosome enrichment at developmental loci is modest, with high significance provided by a moderate average increase at a large number of loci. Histones are also significantly enriched at the promoters of microRNAs (miRNAs) ($P < 0.05$; Supplementary Fig. 6) and at the class of imprinted genes ($P < 0.0001$; Fig. 2), addressed in detail later. Selected loci were tested and confirmed by quantitative PCR (qPCR; Supplementary Fig. 7a–e). Outside of these enriched regions, we observe sequencing reads at low levels distributed genome-wide (for example, Figs 1a and 2a), an observation consistent with low levels of nucleosomes genome-wide, although contributions from non-nucleosomal contamination cannot be ruled out.

¹Howard Hughes Medical Institute, Department of Oncological Sciences, and Huntsman Cancer Institute, ²IVF and Andrology Laboratories, Departments of Surgery, Obstetrics and Gynecology, and Physiology, ³Research Informatics and Bioinformatics Core Facility, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, Utah 84112, USA.

Box 1 | Developmental genes are associated with particular chromatin attributes in human sperm

GoMiner was used to identify enriched categories, and all categories displayed have an FDR < 0.01. The top five general categories are listed, after omitting nearly identical/redundant classes. An expanded gene ontology table with the unfiltered top 30–60 categories, the total genes, number of changed genes, enrichment, and FDR are provided in the Supplementary Information.

Nucleosomes, Array D1

(1) Sequence-specific DNA binding; (2) multicellular organismal development; (3) regulation of transcription; (4) developmental process; (5) regulation of metabolic process.

Nucleosomes, Illumina GAI pooled donors

(1) Transcription factor activity; (2) cell fate commitment; (3) WNT receptor signalling; (4) neuron development; (5) embryonic development.

H3K4me2, Array D1

(1) Multicellular organismal development; (2) developmental process; (3) sequence-specific DNA binding; (4) anatomical structure development; (5) system development.

H3K4me3, Array D1

(1) mRNA processing; (2) RNA binding; (3) cell cycle; (4) transcription; (5) RNA splicing.

H3K4me3, Illumina GAI pooled donors

(1) RNA splicing; (2) translation; (3) cell cycle; (4) RNA metabolic process; (5) transcription.

H3K27me3, Illumina GAI pooled donors

(1) WNT receptor signalling; (2) embryonic organ development and morphogenesis; (3) cell fate commitment; (4) neuron differentiation; (5) sequence-specific DNA binding.

DNA hypomethylated promoters D1 and D2

(1) Embryonic development; (2) multicellular organismal development; (3) system development; (4) RNA biosynthetic process; (5) transcription factor activity.

DNA methylated promoters omitting CpG islands, array

(1) Transcription; (2) RNA biosynthetic process; (3) regulation of transcription; (4) embryonic development; (5) embryo morphogenesis.

important for sperm biology, capacitation and fertilization (Supplementary Table 4), but not at developmental categories. ChIP sequencing (ChIP-seq) analysis with H2A.Z nucleosomes (at standard conditions, 150–250 mM salt) did not show significant enriched Gene Ontology categories, with high enrichment limited to pericentric heterochromatin (Supplementary Fig. 8), consistent with prior immunostaining⁹.

Modified nucleosomes were localized by performing ChIP on mononucleosomes, followed by either array analysis or sequencing (Supplementary Fig. 2, schematic). We normalized the data set for each modification to the data set derived from input mononucleosomes, determined enriched regions (array >towfold; sequencing FDR 40), found the nearest neighbouring gene, and performed Gene Ontology analysis. In somatic cells, H3K4me2 is correlated with euchromatic regions. In sperm, H3K4me2 was enriched at many promoters, and at significant levels at promoters for developmental transcription factors (two replicas $R^2 = 0.94$; Gene Ontology term FDR < 0.06; Box 1 and Supplementary Table 5). In somatic cells, H3K4me3 is localized to: (1) the transcription start sites (TSS) of active genes, (2) genes bearing ‘poised’ RNA polymerase II (Pol II), and (3) the proximal promoter of inactive developmental regulators in embryonic stem (ES) cells—promoters that also bear the silencing mark H3K27me3 (refs 10, 11), and thus termed bivalent. Mature sperm are transcriptionally inert, and Pol II protein levels are barely detectable (data not shown), so the high H3K4me3 levels we observed in sperm chromatin (Supplementary Fig. 1f) seemed surprising. H3K4me3 was localized by both ChIP-chip (three replicas, $R^2 = 0.96$) and ChIP-seq. The raw data sets were similar ($r = 0.7$) and the thresholded data sets were very similar (array twofold; sequencing, FDR 40; 96% intersection, $P < 0.001$). With both data sets, simple inspection showed small peaks at many 5' gene ends, with high levels and broader blocks at a subset of genes (that is, HOX loci; Fig. 1 and Supplementary Fig. 3). Gene Ontology term analyses with either data set yielded genes that are important for changing nuclear architecture, RNA metabolism, spermatogenesis, and also selected transcription factors important for embryonic development (FDR < 0.01, Box 1, Supplementary Tables 6 and 7 and Supplementary Fig. 9). H3K4me3 at genes related to nuclear architecture and spermatogenesis can presumably be attributed to their prior activation during gametogenesis. RNA metabolism occurs both in gametogenesis and the early embryo, so attribution to a prior program as opposed to a potential poising for a future program cannot be unambiguously attributed. However, several transcription and signalling factors of importance in embryo development exhibited high levels and a broad distribution of H3K4me3, including EVX1/2, ID1, STAT3, KLF5, FGF9, SOX7/9, certain HOX genes, and certain noncoding RNAs (Fig. 1 and Supplementary Figs 3 and 6).

Interestingly, ChIP-seq analysis showed significant levels of H3K27me3 at developmental promoters in sperm (Box 1, Fig. 1b, Supplementary Table 8 and Supplementary Figs 3 and 4), and overlapped significantly with H3K27me3-occupied genes in ES cells ($P < 0.01$), which are silent before differentiation. Furthermore, bivalent genes (bearing H3K4me3 and H3K27me3) in ES cells had a significant overlap with bivalent genes in sperm (FDR < 0.001 for each mark). Of the 1,999 genes identified as bivalent in ES cells, 861 were bivalent in sperm ($P < 0.01$; Supplementary Table 9). Also notable but not explored further were many blocks of high H3K4me3 or H3K27me3 in regions lacking annotation (Fig. 1a, oval). Furthermore, H3K9me3 was not detected at the small set developmental promoters tested, but was high at pericentric regions (qPCR only, Supplementary Fig. 7d). Taken together, our results demonstrate extensive histone modification patterns in sperm, and significant similarities to patterns observed in ES cells.

DNA methylation profiles

DNA methylation profiles examined two fertile donors (D2 and D4) using a methylated DNA immunoprecipitation (MeDIP) procedure

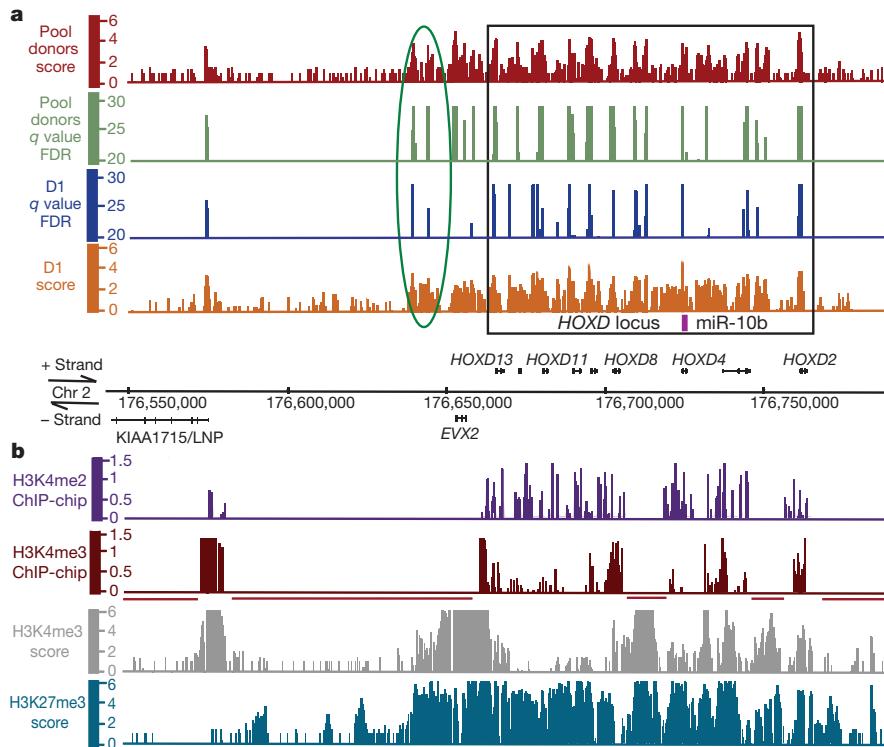


Figure 1 | Profiling of nucleosomes and their modifications at HOXD. For high-throughput sequencing, we show the mapped sequencing reads from D1 or a donor pool (red or orange bars, respectively; normalized difference score), and their significance (green or blue bars; FDR of 20 is <1% and FDR of 30 is <0.1%). **a**, The HOXD locus (black box) and an uncharacterized

flanking locus (green oval). **b**, Profiling of nucleosome modifications at HOXD (in part **a**). The y axis is signal intensity (\log_2 , for ChIP-chip), or the normalized difference score for sequencing. The regions not tiled on the array are underlined in red. Chr, chromosome.

and promoter arrays (individual replicates average $D2 R^2 = 0.97$ and $D4 R^2 = 0.89$). Their methylation patterns were highly similar (pairwise $R^2 = 0.86$), and extensive qPCR validated our array threshold (Supplementary Fig. 7e). Gene Ontology analysis of genes with pronounced DNA hypomethylation yielded transcription and signalling factors that guide embryo development (FDR < 0.05; Box 1 and Supplementary Table 10) including HOX loci (Fig. 3, blue bars, and Supplementary Figs 4 and 10). Hypomethylation also overlapped very significantly with histone-enriched promoters ($P < 0.02$; Supplementary Table 11). Bisulphite sequencing verified the MeDIP results, revealing extensive hypomethylation at developmental promoters in sperm (Supplementary Fig. 10b, c).

Notably, DNA-hypomethylated promoters in mature sperm overlap greatly with developmental promoters bound by the self-renewal network of transcription factors in human ES cells (for example, OCT4 (also known as POU5F1), SOX2, NANOG, KLF4 and FOXD3 proteins¹²; intersection of OCT4 protein occupancy and DNA hypomethylation, $P < 0.01$). In ES cells, these proteins promote self-renewal and also work with repressive polycomb complexes (PRC2; containing core component SUZ12) to help repress a large set of developmental regulators (including HOX genes) to prevent differentiation^{10,13–20}. However, the hypomethylation of developmental genes in sperm is extensive (Fig. 3 and Supplementary Fig. 4). In fact, when CpG islands are omitted from the data sets, Gene Ontology term analysis of hypomethylated promoters still yields developmental genes (Box 1 and Supplementary Table 12). Notably, many of these developmental genes become methylated after differentiation; differential analysis of sperm and primary human fibroblasts (MeDIP, two replicas $R^2 = 0.86$) showed that many promoters occupied by PRC2 in human ES cells acquire methylation in fibroblasts (FDR < 0.01, Supplementary Tables 13 and 14; HOXD illustrated in Fig. 3, Supplementary Figs 4 and 5). Furthermore, the promoters driving several key members of the self-renewal network are themselves markedly hypermethylated in sperm

(OCT4, NANOG and FOXD3, bisulphite sequencing in Supplementary Fig. 10c), whereas their developmental target genes are hypomethylated (bisulphite sequencing in Supplementary Fig. 10b), consistent with recent studies in mice^{21–24}.

Attributes of HOX clusters and miRNAs

Nucleosome enrichment was clear across HOX loci and proximal flanking regions, but falls off precipitously outside (HOXD, Fig. 1a; HOXA, Supplementary Fig. 3a). Histone-enriched HOXD regions with a single donor (D1) were largely shared with the donor pool (Fig. 1a; D1 versus donor pool, $r = 0.7$). Notably, retained nucleosomes have regional covalent modifications. For example, distinct and very large (5–20 kb) blocks of H3K4me3 are clearly observed at all HOX loci, and also at certain imprinted genes (addressed later). At HOXD, high H3K4me3 extends for ~20 kb, encompassing all of EVX2 and extending to the 3' region of HOXD13 (Fig. 1b). Remarkably, a similar profile is observed at the related HOXA locus (Supplementary Fig. 3a). At HOXD a second block of H3K4me3 is observed in the region between HOXD4 and HOXD8 (Fig. 1b), a region that encodes several noncoding RNAs expressed during development. This region represents a marked difference from the chromatin status in ES cells; in ES cells HOXD8–D11 are all bivalent. The distribution of H3K4me2 (determined from two replicas of D1) is clearly different from H3K4me3 at HOX loci (Fig. 1b and Supplementary Fig. 3). For example, at HOXD, H3K4me2 is enriched in HOXD8–D11, a region deficient in H3K4me3 (Fig. 1b). Notably, high H3K27me3 encompasses all HOX loci and their proximal flanking regions. In contrast, high levels of H3K9me (a mark of heterochromatin; Supplementary Fig. 7d) or H2A.Z were not detected at the HOX loci tested.

Histones are enriched at many miRNAs, especially miRNA clusters (Supplementary Fig. 6). For example, 16 of the 29 miRNA clusters on autosomes were significantly enriched ($P < 0.05$). Clusters include those bearing *let7e*, *mir-17*, *mir-15a*, *mir-96*, *mir-135b* and *mir-10a*/

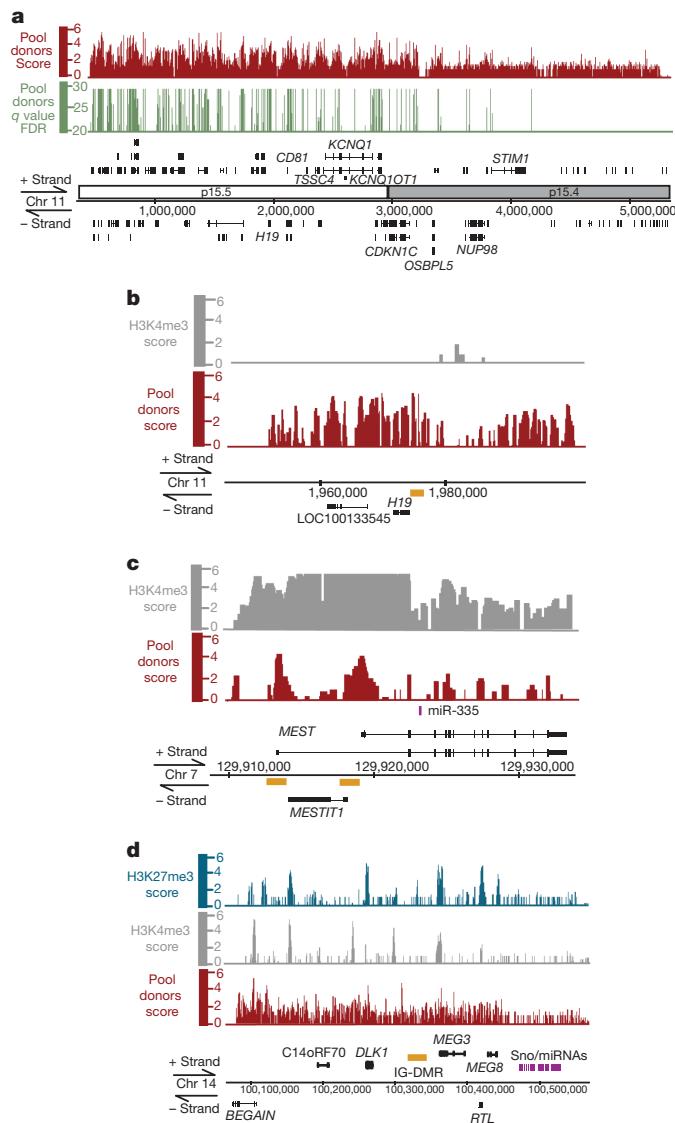


Figure 2 | Nucleosome enrichment at imprinted gene clusters, with high H3K4me3 at paternally expressed noncoding RNAs, and paternally demethylated regions. **a**, Histone enrichment at the 11p15.5 imprinted cluster (ending near *OSBPL5*), but not in the adjacent region. **b, c**, An expanded view of the DMRs (yellow rectangles) of *H19* (paternally methylated) (**b**) and *MEST* (paternally demethylated) (**c**). **d**, Moderate H3K4me3 at the promoters of the paternally expressed genes *BEGAIN*, *DLK1* and *RTL*, and the lack of H3K4me3 at the methylated intergenic-differentially methylated region (IG-DMR) of *MEG3* in sperm. Notably, both H3K4me3 and H3K27me3 reside at the promoter of *MEG3*, which later acquires DNA methylation in the embryo. Sno, small nucleolar.

b, as well as the stand-alone miRNAs *mir-153-1*, *mir-488* and *mir-760*. Notably, many histone-occupied miRNAs are associated with embryonic development²⁵ ($P < 0.01$), and their promoters were largely

hypomethylated (Supplementary Fig. 10d). Furthermore, 7 of the 12 miRNAs on autosomes that are occupied by OCT4, NANOG and SOX2 in human ES cells¹⁷ are also significantly occupied by histone (from pooled sequencing data). However, we do not at present understand the logic for their modification status; certain miRNA clusters have high histone and bivalent status, whereas others lack either modification (Supplementary Fig. 6).

Attributes of primary and secondary imprinted genes

Nucleosomes are significantly enriched at most imprinted genes in sperm, but at both paternally and maternally expressed loci. However, we observe marked specificity of H3K4me3 localization, with high and broad levels present at genes and noncoding RNAs that are paternally expressed. Locus 11p15.5 (Fig. 2a) is a large imprinted cluster with *IGF2*, *H19* and *KCNQ1* and several miRNAs. Here, increased levels of histone are present throughout the imprinted region (up to *OSBPL5*), but not in the large adjacent region lacking imprinted genes (Fig. 2a). Notably, the paternally silenced *H19* locus upstream of *KCNQ1* has a methylated DMR (Supplementary Fig. 10a) that lacks H3K4me3 (Fig. 2b). In contrast, *MEST* (a paternally expressed gene) has high H3K4me3 that extends from its promoter and first exon (containing the demethylated differentially methylated region (DMR); Fig. 2c and Supplementary Fig. 10a) through the second exon. The antisense non-coding RNA *MESTIT1* (also paternally expressed) is transcribed from the first intron, and is also very high in H3K4me3 (Fig. 2c). Furthermore, the promoter region of the paternally expressed antisense noncoding RNA *KCNQ1OT1* displays H3K4me3 (Fig. 2a and data not shown), and the DMR is DNA demethylated (Supplementary Fig. 10a). Several other examples of paternally expressed loci with blocks of H3K4me3 are provided in Supplementary Fig. 11, including *PEG3*, the noncoding RNAs *AIRN* (antisense to *IGF2R*) and *GNASAS* (antisense to *GNAS*). In contrast, genes flanking *KCNQ1* that are repressed by the noncoding RNA *KCNQ1OT1* (such as *OSBPL5*, *TSSC4* and *CD81*; Fig. 2a, expanded in Supplementary Fig. 11) contain histone, but lack H3K4me3. Notably, several paternally silenced genes (bearing DNA methylation) bore moderate (2–3-fold) enrichment of H3K9me3, a mark absent at paternally expressed genes (Supplementary Fig. 7d).

The 14q32.33 region (*DLK-DIO3*) is complex and interesting; paternally expressed genes such as *DLK1* and *RTL1* have moderate levels of H3K4me3 in their promoters, and the imprinting control locus (IG-DMR) lacks H3K4me3 (Fig. 2d) and is DNA methylated^{26–28}. Notably, the promoter of *MEG3* (also known as *GTL2*; just downstream of the IG-DMR) lacks DNA methylation in sperm, but acquires DNA methylation in the embryo^{26–28}, termed secondary imprinting. Notably, the *MEG3* promoter region that later acquires DNA methylation initially bears both H3K4me3 and H3K27me3 in sperm; it is bivalent. One interpretation is that for mature sperm and early embryos, H3K4me3 prevents DNA methylation while H3K27me3 promotes silencing, with subsequent H3K4me removal enabling tissue-specific DNA methylation and secondary imprinting. Furthermore, our examination of the X chromosome inactivation centre showed an apparent bivalent status (and DNA hypomethylation) at the TSS of the *XIST* noncoding RNA, but not at *TSIX*,

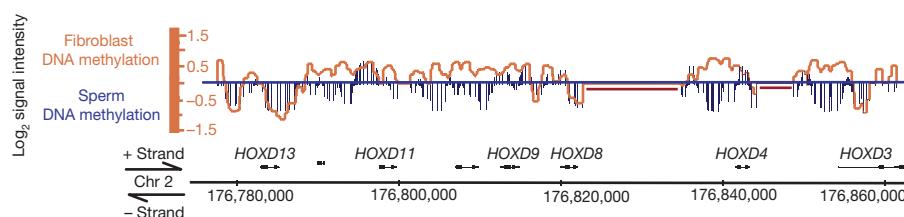


Figure 3 | Developmental promoters in sperm lack DNA methylation, but acquire methylation during development. DNA methylation of the *HOXD* locus in the mature sperm (blue bars) or primary fibroblasts (orange line)

overlay). The y axis is the signal intensity (\log_2) and the x axis is the annotated physical map (HG17). The regions not tiled on the array are underlined in red.

although future studies are required to determine whether these marks influence the regulation of this locus in the embryo (Supplementary Figs 6 and 10d; note that sequence reads on the X chromosome are half that on autosomes, as it is only present in 50% of sperm).

Modifications and expression timing

Transcriptome analysis has been performed in 4-cell and 8-cell human embryos, with 29 or 65 messenger RNAs identified as enriched, respectively²⁹. Notably, genes in sperm bearing H3K4me3 but not H3K27me3 correlated with genes expressed at the 4-cell stage (14 out of 24, $P = 0.059$). Also, genes bearing high H3K4me2 were significantly enriched at genes expressed in the 4–8-cell stage (23 out of 49, $P < 0.02$; only 49 tiled on our array). In contrast, no significant correlation was observed with H3K27me3, which instead associates with transcription factors required for differentiation and organogenesis (discussed earlier). Furthermore, we verified by qPCR the presence of H3K4me2 or H3K4me3 at a subset of these stage-specific gene promoters (Supplementary Fig. 12). Thus, these findings reveal correlations of H3K4me2/3 enrichment, but not H3K27 enrichment, with early expression.

Conclusion

We provide several lines of evidence that the parental genome is packaged and covalently modified in a manner consistent with influencing embryo development. Previous analyses of DNA methylation in sperm identified hypomethylated promoters^{23,24,30,31}, showed similarities to the pattern in ES cells^{24,31}, and overlap between PRC2 and CpG islands^{15,17,21,22}. We add that hypomethylated developmental promoters in human sperm overlap significantly with developmental promoters (in ES cells) occupied by the self-renewal network. Also, the promoters that acquire methylation in fibroblasts are primarily developmental transcription factors that are bound by PRC2 in human ES cells, consistent with recent work linking PRC2 to DNA methylation in development and neuronal differentiation in mice^{21,32,33}. Thus, components of the self-renewal network emerge as candidates for helping to direct DNA hypomethylation in the germ line, and also to guide DNA hypermethylation to particular loci during differentiation, possibly to help ‘lock in’ differentiation decisions, although this remains to be tested.

The central findings of our work involve the significant enrichment of modified nucleosomes in the sperm genome at genes for embryo development, and a specificity to their modification patterns that might be instructive for the regulation of developmental genes, noncoding RNAs and imprinted loci. For example, histone retention and modification were clear at HOX loci and most of the targets of the self-renewal network in ES cells. One key concept in ES cell chromatin is the prevalence of developmental promoters with a bivalent status—bearing both H3K27me3 and H3K4me3 (ref. 10). Many promoters bivalent in ES cells are also bivalent in sperm, although some bear only H3K27me3 in sperm. Notably, H3K27me3 covers essentially all of the four HOX loci in sperm, whereas H3K4me3 is present in large blocks at only a subset of locations in HOX loci. Our work also provides correlations between H3K4me, but not H3K27me, and early expression in the embryo. In contrast, protamine-enriched loci did not show any significant Gene Ontology categories. However, there were certain segments of the Y chromosome with protamine enrichment, including the testis-specific TSPY genes, although the significance is not known.

We also find histones enriched at imprinted gene clusters, and a notable correlation between H3K4me3 and paternally expressed non-coding RNAs and genes; loci that lack DNA methylation in sperm. In contrast, maternally expressed noncoding RNAs/genes, and especially paternally methylated regions, lack H3K4me3 and (for the selected genes tested) contain moderate H3K9me3. Consistent with these observations, recent structural and *in vitro* data show that H3K4 methylation deters DNA methylation by DNMT3A2 and DNMT3L in mice³⁴. However, experiments in model organisms are needed to address whether the modification patterns we report influence

imprinting patterns *in vivo*. Taken together, we reveal chromatin features in sperm that may contribute to totipotency, developmental decisions and imprinting patterns, and open new questions about whether ageing and lifestyle affects chromatin in a manner that impacts fertility or embryo development.

METHODS SUMMARY

Biological samples. Sperm samples were obtained from four men of known fertility attending the University of Utah Andrology laboratory, consented for research. Samples were collected after 2–5 days abstinence and subjected to a density gradient (to purify viable, motile, mature sperm) and treated with somatic cell lysis buffer (0.1% SDS, 0.5% Triton X-100 in DEPC H₂O) for 20 min on ice to eliminate white blood cell contamination. Samples were centrifuged at 10,000g for 3 min, and the sperm pellet was resuspended in PBS and used immediately for chromatin preparation. Clontech human fibroblast cells (Lonza cc-2251) were cultured (37 °C and 5% CO₂) in DMEM containing 10% FBS and supplemented with penicillin and streptomycin.

Chromatin immunoprecipitation. Standard ChIP methods were used³⁵, but we omitted crosslinking and used the following salt concentrations in the numbered buffers³⁵: (1) 150 mM NaCl, (2) 250 mM NaCl, (3) 200 mM LiCl, and (4) 150 mM NaCl (the PBS wash). Antibodies used were: anti-H3K27me3 (Upstate 07-449), H3K4me3 (Abcam 8580), H3K4me2 (Abcam 32356), TH2B (Upstate 07-680), H2A.Z (Abcam 4174) and H3K9me3 (Abcam 8898). For each, 4 µl of antibody was coupled to 100 µl of Dynabeads (Invitrogen). After ChIP, samples for sequencing were not amplified, whereas for arrays the DNA was amplified (WGA, Sigma) before hybridization.

Methylation profiling using MeDIP. MeDIP procedures for sperm and primary human fibroblasts (Clontech) were performed as described previously³⁰.

Sequencing. Sequencing used the Illumina GAII (Illumina Inc.) with standard protocols. Read numbers are final mapped microsatellite filtered reads (26–36 bases). Nucleosomes from D1: 19,658,110, D2–D4: 18,842,467, D1–4: 25,933,196 with equal contribution from each donor (random sub-sampling). Input, human sperm DNA: 17,991,622, H3K4me3: 13,337,105, H3K27me3: 10,344,413, and H2A.Z: 5,449,000. All genomics data sets have been deposited in the Gene Expression Omnibus (GEO) under the SuperSeries GSE15594.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 3 March; accepted 27 May 2009.

Published online 14 June 2009.

1. Ward, W. S. & Coffey, D. S. DNA packaging and organization in mammalian spermatozoa: comparison with somatic cells. *Biol. Reprod.* **44**, 569–574 (1991).
2. Wykes, S. M. & Krawetz, S. A. The structural organization of sperm chromatin. *J. Biol. Chem.* **278**, 29471–29477 (2003).
3. Balhorn, R., Brewer, L. & Corzett, M. DNA condensation by protamine and arginine-rich peptides: analysis of toroid stability using single DNA molecules. *Mol. Reprod. Dev.* **56**, 230–234 (2000).
4. Gatewood, J. M., Cook, G. R., Balhorn, R., Schmid, C. W. & Bradbury, E. M. Isolation of four core histones from human sperm chromatin representing a minor subset of somatic histones. *J. Biol. Chem.* **265**, 20662–20666 (1990).
5. Kimmins, S. & Sassone-Corsi, P. Chromatin remodelling and epigenetic features of germ cells. *Nature* **434**, 583–589 (2005).
6. Nix, D. A., Courdy, S. J. & Boucher, K. M. Empirical methods for controlling false positives and estimating confidence in ChIP-Seq peaks. *BMC Bioinformatics* **9**, 523 (2008).
7. Reik, W., Santos, F. & Dean, W. Mammalian epigenomics: reprogramming the genome for development and therapy. *Theriogenology* **59**, 21–32 (2003).
8. Santos, F., Hendrich, B., Reik, W. & Dean, W. Dynamic reprogramming of DNA methylation in the early mouse embryo. *Dev. Biol.* **241**, 172–182 (2002).
9. Rangasamy, D., Berven, L., Ridgway, P. & Tremethick, D. J. Pericentric heterochromatin becomes enriched with H2A.Z during early mammalian development. *EMBO J.* **22**, 1599–1607 (2003).
10. Bernstein, B. E. et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315–326 (2006).
11. Bernstein, B. E. et al. Methylation of histone H3 Lys 4 in coding regions of active genes. *Proc. Natl. Acad. Sci. USA* **99**, 8695–8700 (2002).
12. Boyer, L. A. et al. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**, 947–956 (2005).
13. Muller, J. & Kassis, J. A. Polycomb response elements and targeting of Polycomb group proteins in *Drosophila*. *Curr. Opin. Genet. Dev.* **16**, 476–484 (2006).
14. Schwartz, Y. B. et al. Genome-wide analysis of Polycomb targets in *Drosophila melanogaster*. *Nature Genet.* **38**, 700–705 (2006).
15. Tanay, A., O'Donnell, A. H., Damelin, M. & Bestor, T. H. Hyperconserved CpG domains underlie Polycomb-binding sites. *Proc. Natl. Acad. Sci. USA* **104**, 5521–5526 (2007).

16. Cao, R. et al. Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298**, 1039–1043 (2002).
17. Lee, T. I. et al. Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* **125**, 301–313 (2006).
18. Takahashi, K. et al. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872 (2007).
19. Kopp, J. L., Ormsbee, B. D., Desler, M. & Rizzino, A. Small increases in the level of Sox2 trigger the differentiation of mouse embryonic stem cells. *Stem Cells* **26**, 903–911 (2008).
20. Wernig, M. et al. *In vitro* reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature* **448**, 318–324 (2007).
21. Mohn, F. et al. Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Mol. Cell* **30**, 755–766 (2008).
22. Illingworth, R. et al. A novel CpG island set identifies tissue-specific methylation at developmental gene loci. *PLoS Biol.* **6**, e22 (2008).
23. Down, T. A. et al. A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. *Nature Biotechnol.* **26**, 779–785 (2008).
24. Farthing, C. R. et al. Global mapping of DNA methylation in mouse promoters reveals epigenetic reprogramming of pluripotency genes. *PLoS Genet.* **4**, e1000116 (2008).
25. Landgraf, P. et al. A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* **129**, 1401–1414 (2007).
26. Glazov, E. A., McWilliam, S., Barris, W. C. & Dalrymple, B. P. Origin, evolution, and biological role of miRNA cluster in DLK-DIO3 genomic region in placental mammals. *Mol. Biol. Evol.* **25**, 939–948 (2008).
27. Takada, S. et al. *Delta-like* and *Gtl2* are reciprocally expressed, differentially methylated linked imprinted genes on mouse chromosome 12. *Curr. Biol.* **10**, 1135–1138 (2000).
28. da Rocha, S. T., Edwards, C. A., Ito, M., Ogata, T. & Ferguson-Smith, A. C. Genomic imprinting at the mammalian *Dlk1-Dio3* domain. *Trends Genet.* **24**, 306–316 (2008).
29. Li, S. S., Liu, Y. H., Tseng, C. N. & Singh, S. Analysis of gene expression in single human oocytes and preimplantation embryos. *Biochem. Biophys. Res. Commun.* **340**, 48–53 (2006).
30. Weber, M. et al. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nature Genet.* **39**, 457–466 (2007).
31. Fouse, S. D. et al. Promoter CpG methylation contributes to ES cell gene regulation in parallel with Oct4/Nanog, Pcg complex, and histone H3 K4/K27 trimethylation. *Cell Stem Cell* **2**, 160–169 (2008).
32. Vire, E. et al. The Polycomb group protein EZH2 directly controls DNA methylation. *Nature* **439**, 871–874 (2006).
33. Schlesinger, Y. et al. Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. *Nature Genet.* **39**, 232–236 (2007).
34. Ooi, S. K. et al. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* **448**, 714–717 (2007).
35. Gordon, M. et al. Genome-wide dynamics of SAPPHIRE, an essential complex for gene activation and chromatin boundaries. *Mol. Cell. Biol.* **27**, 4058–4069 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank B. Dalley for microarray and sequencing expertise, B. Schackmann for oligonucleotides, K. Boucher for statistical analysis, J. Wittmeyer for yeast nucleosomes and helpful comments, and T. Parnell for helpful comments. Financial support was from the Department of Urology (genomics and support of S.S.H.), the Howard Hughes Medical Institute (HHMI) (genomics, biologicals and support of J.P. and H.Z.), CA24014 and CA16056 for core facilities, and the Huntsman Cancer Institute (bioinformatics and support of D.A.N.). B.R.C. is an investigator with the HHMI.

Author Contributions B.R.C., D.T.C. and S.S.H. were involved in the overall design. D.T.C. and S.S.H. were responsible for acquisition of samples, clinical logistics, patient consenting and Institutional Review Board documents. B.R.C., S.S.H., D.A.N. and H.Z. designed detailed molecular and genomics approaches. D.A.N. carried out data processing and array analysis. S.S.H. and D.A.N. performed sequencing analysis. S.S.H. carried out experiments and produced the figures. J.P. carried out immunoblotting and bisulphite sequencing. B.R.C. wrote the manuscript.

Author Information The raw unfiltered reads (fastq format) are deposited at the Gene Expression Omnibus (GEO) under the SuperSeries GSE15594, which encompasses the Subseries entries GSE15690 for ChIP-seq data and GSE15701 for ChIP-chip data. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to D.T.C. (douglas.carrell@hsc.utah.edu) or B.R.C. (brad.cairns@hci.utah.edu).

METHODS

Partitioning of histone- and protamine-associated DNA. Chromatin was prepared from 40 million sperm as described previously³⁶ in the absence of crosslinking reagent, treated with sequential and increasing MNase (10–160 U), and centrifuged to sediment protamine-associated DNA, releasing mononucleosomes. The pooled mononucleosomes were used for ChIP, or the DNA was extracted and gel purified (~140–155 bp) for sequencing and array analysis.

ChIP and preparation for genomics methods. All ChIPs for sequencing were performed using the same pool of mononucleosomes from pooled donors. For arrays, a single pool was used from D1. ChIP methods were as described previously³⁵ but were performed without a crosslinking agent and slight modifications to the salt levels (250 mM NaCl, 200 mM LiCl), and the TE wash was replaced with a 150 mM PBS wash. ChIP methods used anti-H3K27me3 (Upstate 07-449), H3K4me3 (Abcam 8580), H3K4me2 (Abcam 32356), TH2B (Upstate 07-680), or H2A.Z (Abcam 4174) antibodies. For each, 4 µl of antibody was coupled to 100 µl of Dynabeads (Invitrogen). After the ChIP procedure, the DNA was amplified (WGA, Sigma) before hybridization to arrays, whereas samples used for Solexa were not amplified. For sequencing, DNA lengths corresponding to mononucleosomes with adapters (220–280 bp) were gel purified after the addition of the Illumina adaptors. This size selection was also performed for the nucleosomal DNA from pooled donors not subjected to ChIP.

Methylation profiling using MeDIP. This procedure was described previously³⁰. In brief, sonicated sperm DNA was obtained from two different donors and sonicated fibroblast DNA was obtained from Clontech primary human fibroblasts (Lonza CC-2251) (4 µg, 300–1,000-bp fragments). Immunoprecipitated DNA was washed, subjected to whole genome amplification (Sigma Aldrich). Amplified DNA (6 µg) was labelled with Cy5, and input DNA (6 µg) was labelled with Cy3 (Bio labs) by standard methods. Samples were hybridized to Agilent expanded promoter arrays, treated according to standard Agilent conditions, and scanned in an Agilent scanner.

Computational analytical methods. The software used in this analysis are open source and available from the TIMAT2 (<http://timat2.sourceforge.net>) and USeq (<http://useq.sourceforge.net>) project websites. Human annotation and genomic sequence (May 2004, NCBI Build 35, HG17 and March 2006, NCBI Build 36.1, HG18) were obtained from the UCSC Genome Bioinformatic website.

Low-level ChIP-chip analysis. Processing of the Agilent microarray promoter data was performed in three basic steps: data normalization, sliding window summaries, and enriched region identification. For each data set, the median unadjusted signal intensities from the Cy3 and Cy5 channels were extracted. Probes were then mapped to the HG17 or HG18 builds. Biological replicates were quantile normalized and median scaled to 100 (ref. 37). This normalization was applied to the treatment (ChIP samples) and control (whole genomic input DNA for the MeDIP and protamine data sets or DNA derived from mononucleosomes) replicas separately (see later for replica-averaged R^2). Probe level ‘Oligo’ summaries were calculated by taking the \log_2 ratio (mean treatment replicas/mean control replicas). ‘Window’ level summaries were generated by identifying windows of a particular size (100 bp for data sets derived from mononucleosomes, 675 bp for MeDIP and protamine data sets) containing a minimum number of oligonucleotide start positions (one for the data sets derived from mononucleosomes, three for the MeDIP and protamine data sets), and calculating an all pair (treatment versus control) relative difference pseudo median. This window summary score was assigned to the centre position of the window ‘Pse’ or represented as heat map ‘PseHM’ data. Extended regions of high-scoring windows, called ‘intervals’, were identified by merging windows that exceed a set threshold and are located within 250 bp of one another. Intervals were then ranked by their best window score. Relative difference pseudo median scores were converted to \log_2 ratio values.

The average R^2 values for microarray data were as follows: 0.85 for the three D1 MNase replicas; 0.89 for the three Protamine replicas; 0.96 for the two H3C replicas; 0.94 for the two H3K4me2 replicas; 0.93 for the two TH2B replicas; 0.96 for the three H3K4me3 replicas; and 0.93 for the two H3K27me3 replicas. The average MeDIP R^2 values for the three replicas of each donor were as follows: D2 average $R^2 = 0.97$ and D4 = 0.89, and the correlation between D2 versus D4 was 0.87. The average R^2 for the two primary human fibroblast MeDIP replicas was 0.86.

Low-level Chip-seq analysis. The DNA samples derived from mononucleosomes, and the sonicated control input genomic DNA were prepared for sequencing using Illumina’s ChIP-seq kit. The 26-bp and 36-bp reads were generated using Illumina’s Genome Analyser II and their standard software pipeline. Reads were mapped to the March 2006 NCBI Build 36.1 human genome using the pipeline’s eland_extended aligner.

The USeq package⁶ was used to identify regions of histone enrichment relative to input control. This entailed selecting reads that mapped with an alignment score ≥ 13 ($-10\log_{10}(0.05)$), shifting their centre position 73 bp 3' to accommodate the 146-bp mononucleosome fragment length, and using a sliding window of 300 bp to score each region in the genome for significant histone enrichment. Significance was determined by calculating a binomial P value for each 300-bp window and controlled for multiple testing by applying Storey’s q value FDR estimation^{38,39}.

Read numbers. Note the sperm genome has only 4% of the genome in nucleosomes. For nucleosome enrichment D1 had 19,658,110 reads, and the pool of three additional donors had 18,842,467 reads. The raw correlation for D1 versus the donor pool was $r = 0.7$. For all the analysis containing pool donors (D1, and a pooled sample of three additional individuals D2, D3 and D4) we used 25,933,196 mapped filtered reads with equal contribution from each donor (random subsampling). A total of 17,991,622 reads were generated from control input human sperm DNA, 3,337,105 reads from the H3K4me3 sample, 10,344,413 reads for H3K27me3, and 5,449,000 reads for H2Az. The raw unfiltered reads (fastq format) are deposited at GEO under the superseries GSE15594, which encompasses the Subseries entries GSE15690 for ChIP-seq and GSE15701 for ChIP-chip data.

To assess histone enrichment consistency, the QCSeqs application in the USeq package⁶ was used to correlate the read counts between the D1 and pooled sample by calculating a Pearson correlation on the basis of the number of mapped reads falling within 500-bp windowed regions stepped every 250 bp across all chromosomes. Only windows with five or more reads in either of the samples were included in the correlation.

To create lists of candidate histone enriched regions, q -value thresholds of 20 (0.01) and 30 (0.001) ($-10\log_{10}(q \text{ value})$) were selected. Overlapping windows that pass a given threshold were merged and scores from the best window assigned to the enriched region. The normalized window score was then used to rank and sort the regions.

A modification was made to score gene promoters and miRNAs for significant histone enrichment. The first step was to define regions for scoring. For gene promoters, the start of the first exon was used to define its hypothetical promoter by selecting a region 9 kb upstream and 2 kb downstream. For miRNAs, the centre position of each was expanded ± 300 bp. These defined regions were scored for significant enrichment using the window statistics above.

High-level ChIP-chip and ChIP-seq analysis. Intersect regions. To identify regions of significant intersection between enriched region lists from various data sets, the USeq IntersectRegions application was used. This application counts the number of intersections between two lists of genomic coordinates that occur within a minimum ‘max gap’ distance. To estimate confidence in the intersections, a thousand ‘random’ data sets are generated that were matched to the chromosome and size of the original regions, and randomly picked from the interrogated regions on the array or sequenced regions in the genome. These randomized data sets were used to calculate a P value for the intersection and fold enrichment (fraction real intersection/fraction average random data set intersection) over random. Initial pilots that imposed a fraction GC match when picking random regions showed little difference with non-GC-matched random data sets and were thus subsequently dropped.

Find neighbouring genes (FNG). Genes associating with histones or histone modifications were determined using the FNG application in the USeq package. The gene lists were uploaded in GoMiner (<http://discover.nci.nih.gov/gominer/> htgm.jsp) to identify over represented Gene Ontology terms.

Intersect lists. To determine whether the 4- and 8-cell transcripts identified in early human embryo correlated with any of our histone modifications we used The IntersectLists USeq application which uses random permutation to calculate the significance of intersection between two lists of genes.

Aggregate plots. The USeq AggregatePlots application was used to compare the degree of enrichment and distribution of histone reads surrounding the TSS of developmental and non developmental genes. The gene classes were derived on the basis of Gene Ontology term categories.

36. Zalenskaya, I. A., Bradbury, E. M. & Zalensky, A. O. Chromatin structure of telomere domain in human sperm. *Biochem. Biophys. Res. Commun.* **279**, 213–218 (2000).
37. Bolstad, B. M., Irizarry, R. A., Astrand, M. & Speed, T. P. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**, 185–193 (2003).
38. Dudoit, S., Gilbert, H. N. & van der Laan, M. J. Resampling-based empirical Bayes multiple testing procedures for controlling generalized tail probability and expected value error rates: focus on the false discovery rate and simulation study. *Biom. J.* **50**, 716–744 (2008).
39. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. USA* **100**, 9440–9445 (2003).

ARTICLES

Argonaute HITS-CLIP decodes microRNA–mRNA interaction maps

Sung Wook Chi¹, Julie B. Zang¹, Aldo Mele¹ & Robert B. Darnell¹

MicroRNAs (miRNAs) have critical roles in the regulation of gene expression; however, as miRNA activity requires base pairing with only 6–8 nucleotides of messenger RNA, predicting target mRNAs is a major challenge. Recently, high-throughput sequencing of RNAs isolated by crosslinking immunoprecipitation (HITS-CLIP) has identified functional protein–RNA interaction sites. Here we use HITS-CLIP to covalently crosslink native argonaute (Ago, also called Eif2c) protein–RNA complexes in mouse brain. This produced two simultaneous data sets—Ago-miRNA and Ago-mRNA binding sites—that were combined with bioinformatic analysis to identify interaction sites between miRNA and target mRNA. We validated genome-wide interaction maps for miR-124, and generated additional maps for the 20 most abundant miRNAs present in P13 mouse brain. Ago HITS-CLIP provides a general platform for exploring the specificity and range of miRNA action *in vivo*, and identifies precise sequences for targeting clinically relevant miRNA–mRNA interactions.

Sophisticated mechanisms regulating RNA may explain the gap between the great complexity of cellular functions and the limited number of primary transcripts. Regulation by miRNAs underscores this possibility, as each miRNA is believed to bind directly to many mRNAs to regulate their translation or stability^{1,2}, and thereby control a wide range of activities, including development, immune function and neuronal biology^{3–5}. Many miRNAs are evolutionarily conserved, although others are species-specific (including human miRNAs not conserved in chimpanzee)⁶, consistent with roles ranging from generating cellular to organismal diversity.

Despite their biological importance, determining the targets of miRNAs is a major challenge. The problem stems from the discovery that functional mRNA regulation requires interaction with as few as 6 nucleotides of miRNA seed sequence⁷. Such 6-mers are present on average every ~4 kilobases (kb), so that miRNAs could regulate a broad range of targets; however, the full extent of their action is not known. Bioinformatic analysis has greatly improved the ability to predict bona fide miRNA binding sites^{8–10}, principally by constraining searches for evolutionarily conserved seed matches in 3' untranslated region (UTR). Nonetheless, different algorithms produce divergent results with high false-positive rates^{3,10–12}. In addition, many miRNAs are present in closely related miRNA families, complicating interpretation of loss-of-function studies in mammals^{13,14}, although such studies have been informative for several miRNAs^{3,15–17}. miRNA over-expression or knockdown studies, most recently in combination with proteomic studies^{11,18}, have led to the conclusion that individual miRNAs generally regulate a relatively small number of proteins at modest levels (<2-fold), although the false-positive rate of target predictions remains high (~up to 66%)¹¹, and the data sets analysed have been of limited size (~5,000 proteins). Similar high false-positive rates have been observed when miRNAs were co-immunoprecipitated with Ago proteins^{19–23}. A critical caveat common to all of these studies is their inability to definitively distinguish direct from indirect miRNA–target interactions. At the same time, as therapeutic antisense strategies become more viable^{17,24,25}, knowledge of direct miRNA target sites has become increasingly important.

Recently, we developed HITS-CLIP to directly identify protein–RNA interactions in living tissues in a genome-wide manner^{26,27}. This

method^{28,29} uses ultraviolet irradiation to covalently crosslink RNA–protein complexes that are in direct contact (approximately over single ångstrom distances) within cells, allowing them to be stringently purified. Partial RNA digestion reduces bound RNA to fragments that can be sequenced by high-throughput methods, yielding genome-wide maps and functional insights^{26,30}. Recent X-ray crystal structures of an Ago–miRNA–mRNA ternary complex³¹ suggest that Ago may make sufficiently close contacts to allow Ago HITS-CLIP to simultaneously identify Ago-bound miRNAs and the nearby mRNA sites. Here we use Ago HITS-CLIP to define the sites of Ago interaction *in vivo*, decoding a precise map of miRNA–mRNA interactions in the mouse brain. This provides a platform that can establish the direct targets on which miRNAs act in a variety of biological contexts, and the rules by which they do so.

Ago RNA targets in the mouse brain

HITS-CLIP experiments rely on a means of purifying RNA-binding proteins (RNABPs)^{26–29}. To purify Ago bound to mouse brain RNAs, we ultraviolet-irradiated P13 neocortex and immunoprecipitated Ago under stringent conditions. After confirming the specificity of Ago immunoprecipitation (Fig. 1a), we radiolabelled RNA, further purified crosslinked Ago–RNA complexes by SDS–polyacrylamide gel electrophoresis and nitrocellulose transfer, and visualized them by autoradiography. We observed complexes of two different modal sizes (~110 kDa and ~130 kDa; Fig. 1b and Supplementary Fig. 1), suggesting that Ago (97 kDa) was crosslinked to two different RNA species. Polymerase chain reaction with reverse transcription (RT–PCR) amplification revealed that the ~110-kDa lower band harboured ~22-nucleotide crosslinked RNAs and the upper band both 22-nucleotide and larger RNAs (Fig. 1c). These products were sequenced with high-throughput methods²⁶ and found to correspond to miRNAs and mRNAs, respectively (Supplementary Table 1), suggesting that Ago might be sufficiently close to both miRNA and target mRNAs to form crosslinks to both molecules in the ternary complex (Fig. 1d). Such a result would allow the search for miRNA binding sites to be constrained to both the subset of miRNAs directly bound by Ago and to the local regions of mRNAs to which Ago crosslinked, potentially reducing the rate of false-positive predictions of miRNA binding.

¹Laboratory of Molecular Neuro-Oncology and Howard Hughes Medical Institute, The Rockefeller University, 1230 York Avenue, New York, New York 10021, USA.

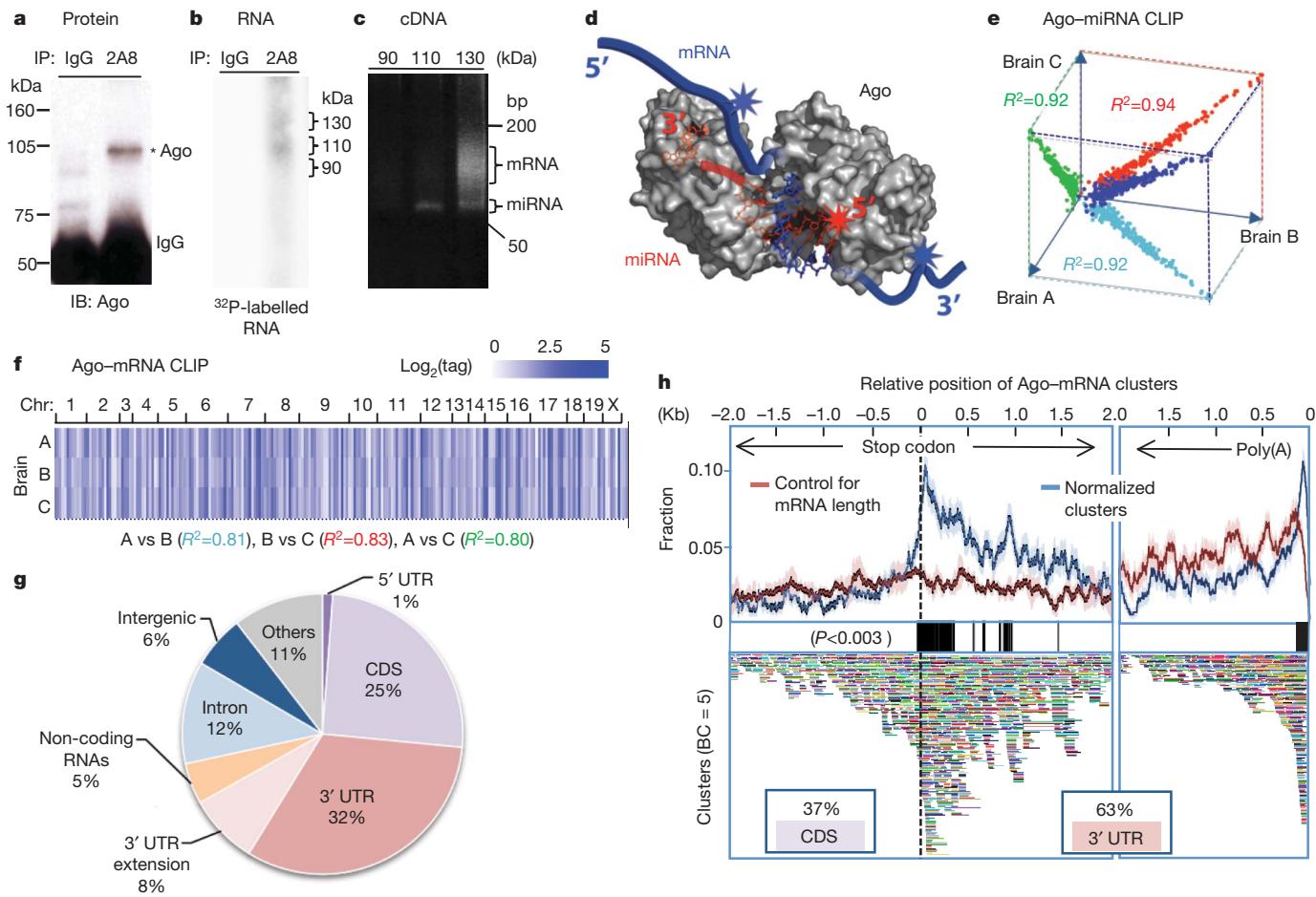


Figure 1 | Argonaute HITS-CLIP. **a**, Immunoblot (IB) analysis of Ago immunoprecipitates (IP) from P13 mouse neocortex using pre-immune IgG as a control or anti-Ago monoclonal antibody 2A8 blotted with 7G1-1* antibody (Supplementary Methods). **b**, Autoradiogram of ^{32}P -labelled RNA crosslinked to mouse brain Ago purified by immunoprecipitation. RNA–protein complexes of ~ 110 kDa and ~ 130 kDa are seen with 2A8 but not control immunoprecipitation. **c**, PCR products amplified after linker (36 nucleotides) ligation to RNA products excised from **b**. Products from the 110-kDa RNA–protein complex were ~ 22 -nucleotide miRNAs, and those from 130-kDa complexes were predominantly mRNAs. **d**, Illustration showing proposed interpretation of data in **c**. Ago (drawn based on structure 3F73 in the Protein Data Bank)³¹ binds in a ternary complex to both miRNA and mRNA, with sufficiently close contacts to allow ultraviolet crosslinking to either RNA–mRNA tags will be in the immediate vicinity of miRNA binding sites. **e**, **f**, Reproducibility of all Ago–miRNA tags (**e**; shown as $\log_2(\text{normalized miRNA frequency})$ per brain) or all tags within Ago–mRNA

To differentiate robust from nonspecific or transient Ago–RNA interactions, we compared the results from biological replicate experiments done with two different monoclonal antibodies (Supplementary Figs 1–3). The background was further reduced by *in silico* random CLIP, a normalization algorithm that accounted for variation in transcript length and abundance (Supplementary Figs 4 and 5). The set of Ago-crosslinked miRNAs and mRNAs was highly reproducible. Among biological triplicates or among 5 replicates done with two antibodies, the Pearson correlation coefficient was $R^2 > 0.9$ and > 0.83 , respectively, for Ago–miRNA CLIP (Fig. 1e and Supplementary Fig. 2) and $R^2 \geq 0.8$ and ≥ 0.65 , respectively, for Ago–mRNA CLIP (Fig. 1f and Supplementary Fig. 3). We identified 454 unique miRNAs crosslinked to Ago in mouse brain, with Ago–miR-30e being the most abundant species (14% of total tags; Supplementary Fig. 2); these results were consistent with previous estimates assessed by cloning frequency³² or bead-based cytometry³³, although the correlation with published results ($R^2 = 0.2$ –0.32;

clusters (**f**; see Supplementary Figs 2 and 3). We estimated that sequencing depth was near saturation (Supplementary Fig. 16). **g**, Location of reproducible Ago–mRNA tags (tags in clusters; $BC \geq 2$) in the genome. Annotations are from RefSeq: ‘others’ are unannotated EST transcripts; non-coding RNAs are from lncRNAs or FANTOM3. **h**, Top panel: the position of robust Ago–mRNA clusters ($BC = 5$) in transcripts is plotted relative to the stop codon and 3' end (presumptive poly(A) site, as indicated). Data are plotted as normalized density relative to transcript abundance for Ago–mRNA clusters (blue) or control clusters (red) (s.d. is shown in light colours; see Supplementary Methods). Regions with significant enrichment relative to control are indicated with black bars (> 3 s.d.; $P < 0.003$). Cluster enrichment ~ 1 kb downstream from the stop codon appears to be due to a large number of transcripts with ~ 1 kb 3' UTRs (data not shown). Bottom panel: all individual clusters ($BC = 5$) are shown (each is a different colour).

Supplementary Fig. 6) was not as high as among our biological replicates. These discrepancies might be due to differences in the ages of brain used, regulation of Ago–mRNA interactions, and/or increased sensitivity allowed by stringent CLIP conditions and consequent improved signal/noise. To facilitate the analysis of large numbers of Ago–mRNA CLIP tags ($\sim 1.5 \times 10^6$ unique tags; Supplementary Table 1) we analysed overlapping tags (clusters)²⁶, which were normalized by *in silico* random CLIP and sorted by biological complexity²⁶ (BC , a measure of reproducibility between biological replicates; see Supplementary Figs 5 and 7). A total of 1,463 robust clusters ($BC = 5$; that is, harbouring CLIP tags in all five biological experiments using both antibodies) mapped to 829 different brain transcripts, and 990 clusters had at least 10 tags (Supplementary Fig. 7).

Ago–mRNA HITS-CLIP tags were enriched in transcribed mRNAs (Fig. 1g). The pattern of tags mirrored the results of functional assays with miRNAs³⁴, which show no biological activity when seed sites are present in 5' UTRs (1% Ago–mRNA tags), and high efficacy in 3'

UTRs (40% tags including 8% within 10 kb downstream of annotated transcripts, regions likely to have unannotated 3' UTRs^{26,35}). In addition, an extensive set of tags was identified in other locations, including coding sequence (25%), a location for which there is emerging evidence of miRNA regulation^{36–40}, introns (12%), and non-coding RNAs (4%), indicating that these sites may provide new insights into miRNA biology. Within mRNAs, Ago–mRNA clusters were highly enriched in the 3' UTR (~60%; Supplementary Fig. 8), especially around stop codons (with a peak ~50 nucleotides downstream) and at the 3' end of transcripts (~70 nucleotides upstream of presumptive poly(A) sites; $P < 0.003$, Fig. 1h), consistent with bioinformatic observations from microarray data³⁴. Taken together, these data suggested that Ago–mRNA clusters might be associated with functional binding sites.

To examine the relationship between Ago–mRNA clusters and potential sites of miRNA action, we first performed an unbiased search for all 6–8-nucleotide sequence motifs within Ago–mRNA clusters using linear regression analysis (Supplementary Methods). The six most enriched motifs corresponded to seed sequences present within the most frequently crosslinked miRNAs in Ago–miRNA CLIP (Ago–miRNAs; Supplementary Table 2). The most significant match corresponded to the seed sequence of miR-124, a well-studied brain-specific miRNA ($P = 8.3 \times 10^{-58}$; because miR-124 is only the eighth most frequently crosslinked Ago–miRNA, these data may indicate over-representation of miR-124 seed matches in the genome (Supplementary Tables 2 and 3), or contributions from currently unknown rules of Ago binding). To define more precisely the region of mRNA complexed with Ago, we examined the width of 61 robust Ago–mRNA clusters ($BC = 5$; total tags >30) relative to their peaks (determined by cubic spline interpolation; Supplementary Methods). We found that Ago bound within 45–62 nucleotides of

cluster peaks $\geq 95\%$ of the time (Fig. 2a), and we defined this region as the average Ago–mRNA footprint.

Within Ago–mRNA footprints (11,118 clusters; $BC \geq 2$) we found a high correlation between the frequency of Ago–miRNAs and the frequency of their seed matches (Fig. 2b and Supplementary Fig. 9). The seed matches were near the peaks of 134 robust clusters (Fig. 2c and Supplementary Table 3), and their distribution was leptokurtic (with a higher and sharper peak than a normal distribution; excess kurtosis (k) = 1.08, versus 0 in a normal distribution). As a control, seed matches for a ‘negative’ group of miRNAs (the lowest ranks from the Ago–miRNA list) were uniformly distributed across these clusters (Fig. 2c; $k = -1.35$, versus -1.2 in uniform distribution). Taken together, these results indicate that the Ago–mRNA footprint is rich in, and may predict, miRNA binding sites with enhanced specificity over other approaches (at least a threefold improvement in false positives; Supplementary Methods).

Ago HITS-CLIP and miR-124

To explore further the relationship between the Ago–mRNA footprint and miRNA binding we focused on miR-124 sites. There was a marked enrichment of conserved miR-124 seed matches in Ago–mRNA clusters ($BC \geq 2$; Fig. 2d, estimated false-positive rate of 13%; Supplementary Methods). Eighty-six per cent of the predicted miR-124 binding sites were present within the Ago footprint region, again in a tight peak region showing leptokurtic distribution (Fig. 2d; $k = 11.63$). Although some predicted seeds outside of the Ago footprint might correspond to false positives, we noted small secondary peaks at $\sim \pm 50$ nucleotides outside the Ago footprint (Fig. 2d), indicating the possibility of cooperative secondary miR-124 binding sites in some transcripts, consistent with previous data³⁴. Relative to more stringent analyses (Fig. 2c), our analysis at this threshold ($BC \geq 2$)

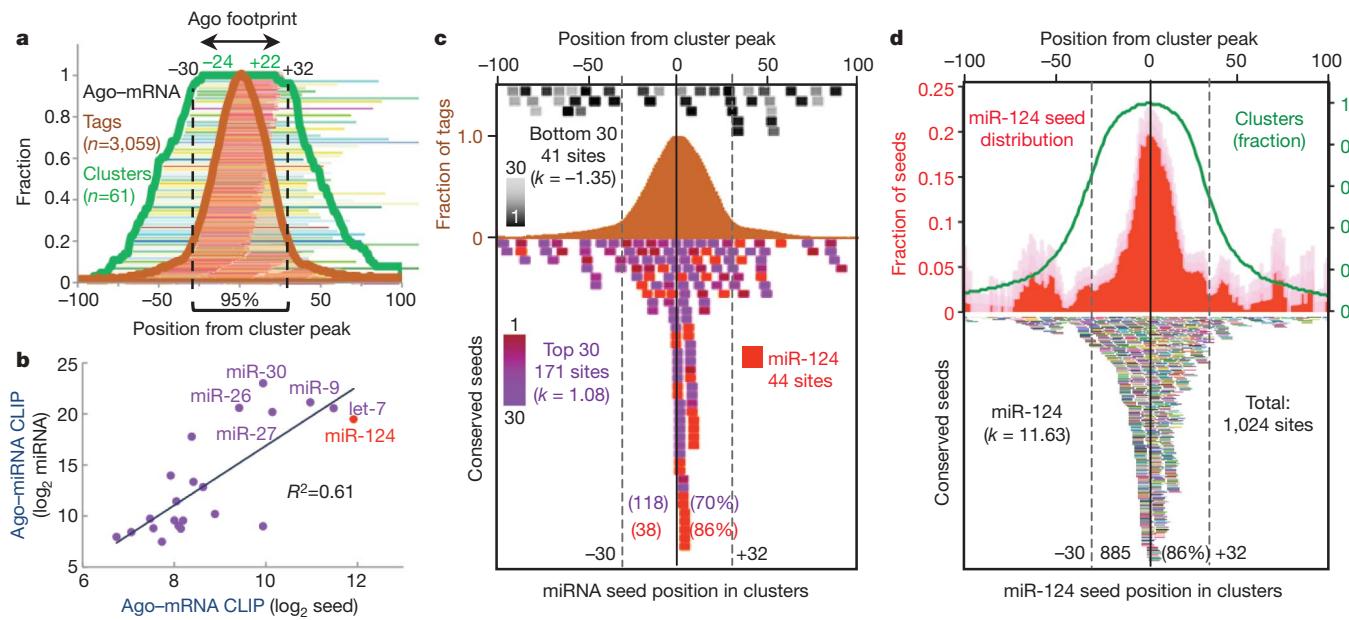


Figure 2 | Distribution of mRNA tags correlates with seed sequences of miRNAs from Ago CLIP. **a**, Ago–mRNA cluster width and peaks. The peaks of 61 robust clusters ($BC = 5$, peak height >30 , with single peaks) were determined, and the position of tags (brown lines and fraction plotted as brown graph) and width of individual clusters (green lines and fraction plotted as green graph) are shown relative to the peaks (Supplementary Methods). The minimum region of overlap of all clusters (100%) was within -24 and $+22$ nucleotides of cluster peaks, and $\geq 95\%$ were within -30 and $+32$ nucleotides, suggesting that the Ago footprint on mRNA spans 62 nucleotides (or, more stringently, 46 nucleotides). **b**, A linear regression model was used to compare miRNA seed matches enriched in the stringent Ago footprint region with the frequency of miRNAs experimentally determined by Ago–miRNA HITS-CLIP (see also Supplementary Fig. 9).

c, The positions of conserved core seed matches (in position 2–7) from the top 30 Ago-bound miRNAs (independent of peak height; purple colours; including miR-124, red; or, bottom 30 miRNAs, expressed at extremely low levels in brain (Supplementary Fig. 2); black to grey colours) are plotted relative to the peak of 134 robust clusters ($BC = 5$, peak height >30). A total of 118 of 171 seed matches are within the Ago 62-nucleotide footprint. **d**, The positions of conserved miR-124 seed matches (bottom panel; each is represented by a different colour) were plotted relative to the peak position of all Ago–mRNA clusters ($BC \geq 2$). Top panel: distribution of miR-124 seed matches (plotted relative to cluster peak, normalized to number of clusters; red graph). The pale colour indicates s.d. Excess kurtosis (k) indicates that seed sites are present in a sharp peak relative to a normal distribution.

was more sensitive and sufficiently specific such that we used it for subsequent analyses (Supplementary Fig. 7).

We searched published examples of miR-124-regulated transcripts for Ago–mRNA clusters harbouring miR-124 seed matches within the predicted 62-nucleotide Ago footprint (referred to as Ago–miR-124 ternary clusters). We identified such ternary clusters in 5 of 5 transcripts in which miR-124 seed sites had been well defined by functional studies (including mutagenesis of seed sequences in full-length 3' UTR; Fig. 3 and Supplementary Fig. 10). In each of these transcripts, there were many predicted miRNA target sites in the 3' UTR relative to the small number of Ago–mRNA ternary clusters found, suggesting that there may be a significant number of false-positive predictions from bioinformatic algorithms (Fig. 3 and Supplementary Fig. 10). For example, the 3' UTR of *Itgb1* mRNA has ~50 predicted miRNA target sites, including two miR-124 sites, but only 5 Ago–mRNA ternary clusters (Fig. 3a). Using the Ago footprint to predict which miRNAs bound at these sites (Ago ternary map; Supplementary Methods) we identified three as miR-124 sites, one of which was not predicted computationally because the seed sequence is not conserved (Fig. 3a); similar observations were made in the *Ctdsp1* 3' UTR (Supplementary Fig. 10). Previous luciferase assays demonstrated that miR-124 suppression of *Itgb1* (to 35% control levels) was partially reversed (to 85% of control levels) by mutating both of the two predicted seed sequences⁴¹; our observation of an Ago–miR-124 ternary cluster at this third non-conserved site may explain the partial rescue. Conversely, in the *Ptbp1* 3' UTR, the absence of any Ago footprint at a predicted miR-9 seed site was consistent with previous studies which found this site to be

non-functional⁴² (Fig. 3b). Additionally, *Ptbp1* has seven predicted miR-124 seed sites, of which five were previously tested and only two found to be functional in luciferase assays⁴²; only these two sites harboured Ago–miR-124 ternary clusters (Fig. 3b).

To extend these observations, we compiled brain-expressed transcripts from a meta-analysis of five published microarray experiments which identified transcripts downregulated after miR-124 overexpression in HeLa and other cell lines (Supplementary Methods)^{7,11,21,42,43}. Brain-expressed transcripts that had predicted 3' UTR miR-124 seeds were also suppressed at the mRNA and protein level by miR-124 overexpression (Fig. 4), consistent with previous experiments^{11,18}. However, transcripts with Ago–miR-124 ternary clusters had a significantly greater tendency to be downregulated at the mRNA and protein level ($P < 0.01$, Kolmogorov–Smirnov test, Fig. 4). Ago–miR-124 ternary clusters had much greater predictive value (true positive rate = 73%) and specificity (92.5%) than analysis of conserved seed sequences alone (Supplementary Fig. 11). We validated these studies experimentally by examining Ago–mRNA clusters that appeared *de novo* in HeLa cells (which do not express significant ($<0.03\%$) miR-124) after miR-124 transfection. Applying these data to the meta-analysis, mRNAs for which 3' UTRs harboured new Ago–miR-124 ternary clusters after miR-124 transfection showed an even greater enrichment in miR-124-dependent changes in transcript (Fig. 4c) and protein (Fig. 4d) levels ($P < 0.01$, Kolmogorov–Smirnov test).

We next examined Ago–miR-124 ternary clusters present in validated individual transcripts identified from among 168 candidate miR-124-regulated transcripts⁷. These targets were analysed previously⁴⁴ using a

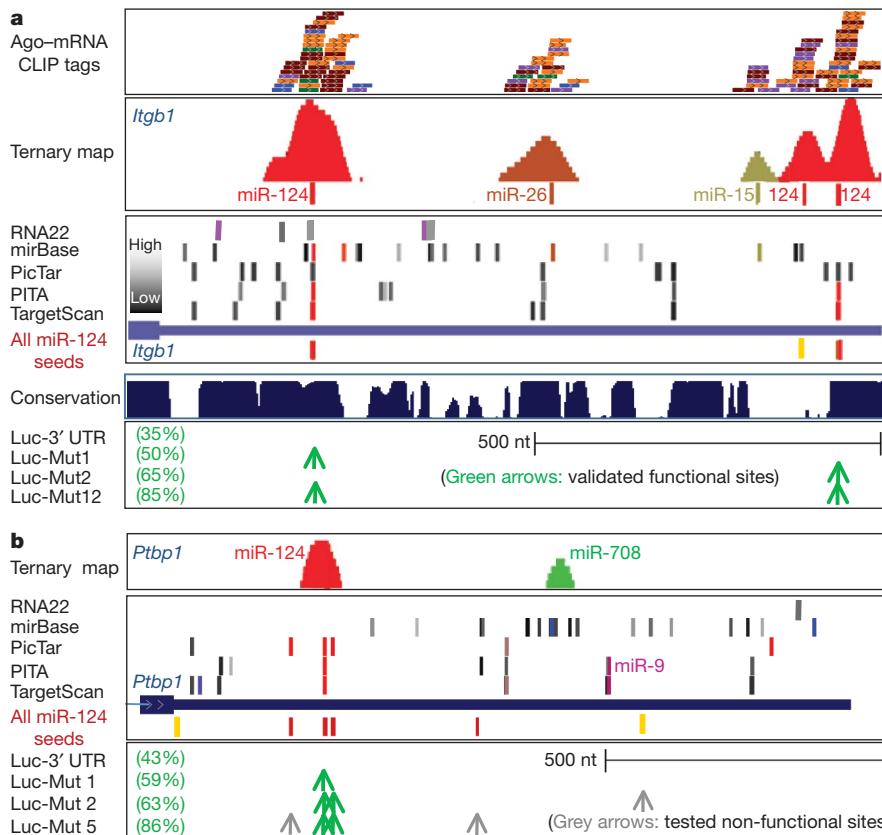


Figure 3 | Ago–miRNA ternary clusters in validated miR-124 mRNA targets. **a**, Ago–mRNA CLIP tags (top panel, raw tags, one colour per biological replicate; second panel, ternary map of Ago–mRNA normalized clusters around top predicted miRNA sites) compared with predicted miRNA sites (using indicated algorithms) for the 3' UTR of *Itgb1* (third panel, see Supplementary Methods; colours indicate predicted top 20 miRNAs as in Fig. 6; grey bars indicate miRNAs ranking below the top 20 (per heat map) in Ago–miRNA CLIP). All predicted miR-124 6–8-mer seeds

(conserved (red) or non-conserved (yellow) sites) are shown. The bottom panel shows data from luciferase assays in which mutagenesis of predicted miR-124 seeds at the indicated positions had the indicated effects on rescuing miR-124-mediated suppression (35% of baseline luciferase levels)⁴¹. **b**, Ago–miRNA ternary maps compared with previously reported functional data for *Ptbp1* (ref. 42) (see Supplementary Fig. 10 for maps of *Ctdsp1* (ref. 44) and *Vamp3* (ref. 44)).

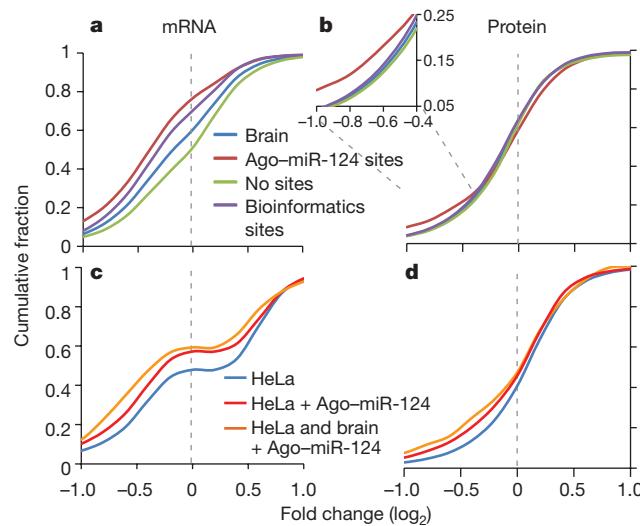


Figure 4 | Meta-analysis of Ago-mRNA clusters in large-scale screens of miR-124-regulated targets. **a**, Transcripts with predicted (conserved) miR-124 seeds (purple line) showed miR-124 suppression relative to all transcripts expressed in brain and cell lines (blue line) or those with no miR-124 seed sequences (green line). Transcripts with Ago-miR-124 ternary clusters (containing both miR-124 seed sequences and Ago-mRNA CLIP tags; red line) showed further miR-124 suppression. **b**, Similar results were seen when analysing miR-124-dependent protein suppression (identified by SILAC)¹¹, with discrimination by the presence of Ago-miR-124 ternary clusters especially evident where there were smaller numbers of transcripts showing larger changes ($\log_2 < -0.4$; inset). **c**, Transcripts expressed in miR-124-transfected HeLa cells that harbour new Ago-miR-124 clusters (red line; or a subset of transcripts also harbouring Ago-miR-124 clusters in mouse brain; yellow line), compared with previous analysis⁷ of regulated transcripts in miR-124-transfected HeLa cells. **d**, As in **c**, plotted for predicted protein levels, compared with previous data¹¹.

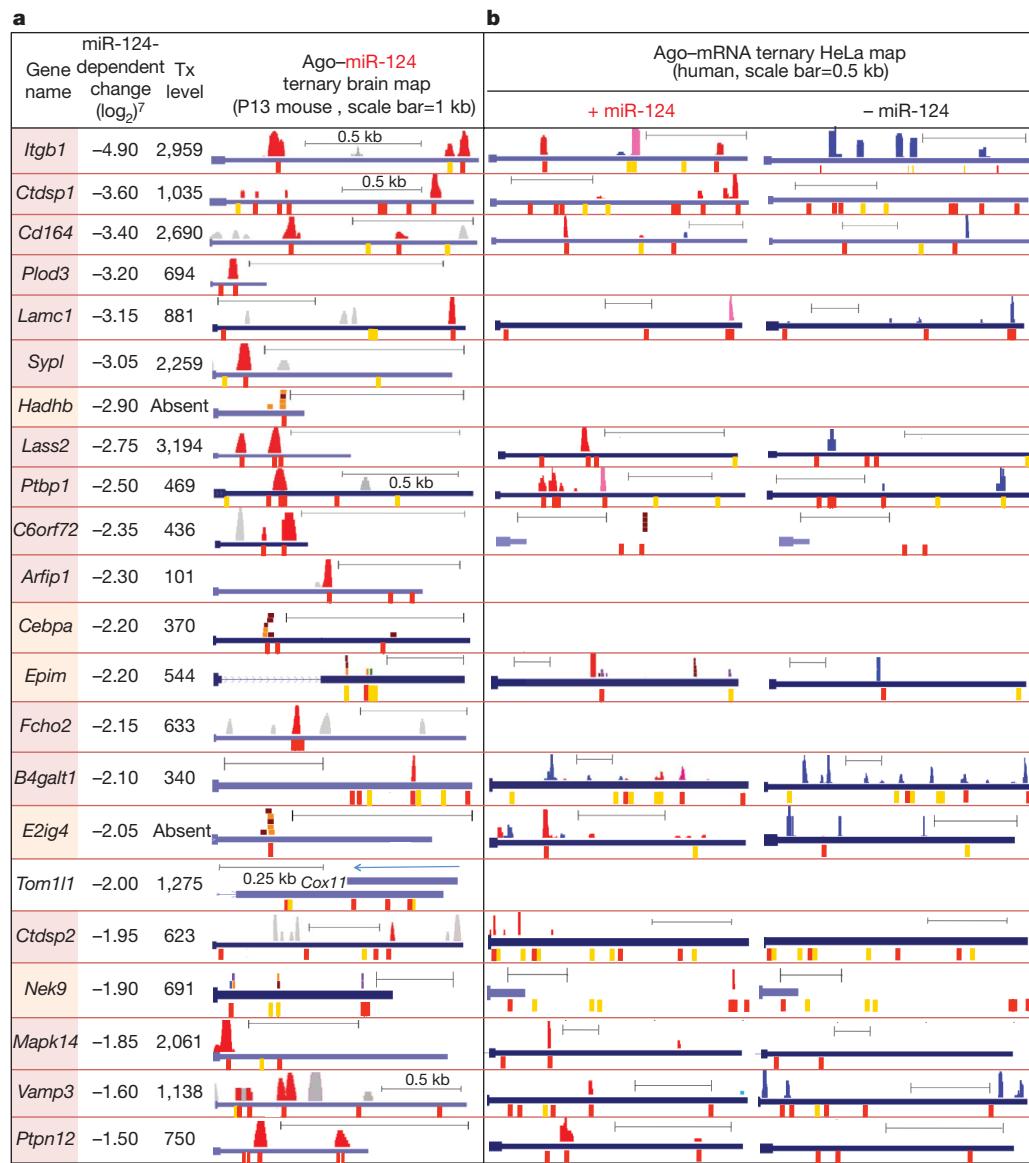


Figure 5 | Ago-miR-124 ternary maps in brain and transfected HeLa cells. **a**, Comparison of 22 validated miR-124 targets⁴⁴ with Ago-miR-124 ternary clusters in brain transcripts (BC ≥ 2); miR-124 clusters are shown graphically above each 3' UTR, as in Fig. 3, and corresponding conserved (red) or non-conserved (yellow) miR-124 seeds (below) were identified in 16 of 22 validated transcripts (gene names shaded red). In 5 additional transcripts (gene names shaded orange) clusters were below normalization threshold, but individual Ago-mRNA CLIP tags are shown. *Tom1l1* had no detectable Ago binding, although notably a brain-expressed antisense transcript (*Cox11*) is present. Previously reported changes in miR-124-dependent transcript levels from miR-124-transfected HeLa cells⁷, and levels of the same transcript ('Tx') in mouse brain (normalized probe intensities), are shown. **b**, Columns represent Ago-miR-124 ternary clusters in HeLa cells transfected with either miR-124 ('+miR-124') or a control miRNA ('-miR-124'). Seventeen *de novo* Ago-miR-124 ternary clusters (red; present only after miR-124 transfection) are shown, together with clusters present in both control and miR-124-transfected cells (purple), or present in control HeLa cells alone (blue).

rigorous three-part strategy to validate experimentally 22 of them (although miR-124 binding sites were not generally defined). Sixteen out of twenty-two harboured Ago–miR-124 ternary clusters in the 3' UTR (Fig. 5), and in five additional transcripts with low expression levels, Ago–mRNA CLIP tags were identified at predicted miR-124 seed sites. For transcripts of even moderate abundance (with normalized probe intensities ≥ 700 ; average for P13 brain transcripts $\sim 1,255$; Supplementary Methods), we identified all ten predicted targets (Fig. 5). These data indicate that identifying Ago–miR-124 ternary clusters markedly enhances the sensitivity and specificity of detecting bona fide miR-124 targets.

We examined this set of 22 targets⁴⁴ for Ago–mRNA clusters that appeared in the 3' UTR after miR-124 transfection (Fig. 5). Notably, from among many potential miR-124 seed sites, 17 *de novo* Ago–miR-124 ternary clusters appeared after transfection and 14 of 17 were at precisely the positions predicted from the brain Ago–miR-124 maps. Genome-wide, we identified 204 *de novo* Ago–miR-124

clusters with conserved seeds in mouse brain transcripts; of these 98 were independently identified as mouse brain Ago–miR-124 ternary clusters. Taken together, these results confirmed that the Ago ternary map identifies functional sites of miRNA regulation.

Predicting miRNA functional networks

On the basis of the robust correlation between previously validated miR-124 functional sites and Ago HITS-CLIP, we examined brain Ago–mRNA clusters to predict binding maps for the 20 most abundant Ago–miRNAs (Supplementary Methods). These maps (Fig. 6a) show that Ago binds to target transcripts at very specific sites: on average there are only 2.6 Ago–mRNA clusters ($BC \geq 2$) per Ago regulated transcript (2.12 per 3' UTR) and each miRNA binds an average of only 655 targets (Supplementary Fig. 14). To explore the potential of Ago HITS-CLIP maps to define miRNA regulated transcripts, we examined the functions encoded by the predicted targets of these 20 miRNAs using gene ontology (GO) analysis: comparison of these results with

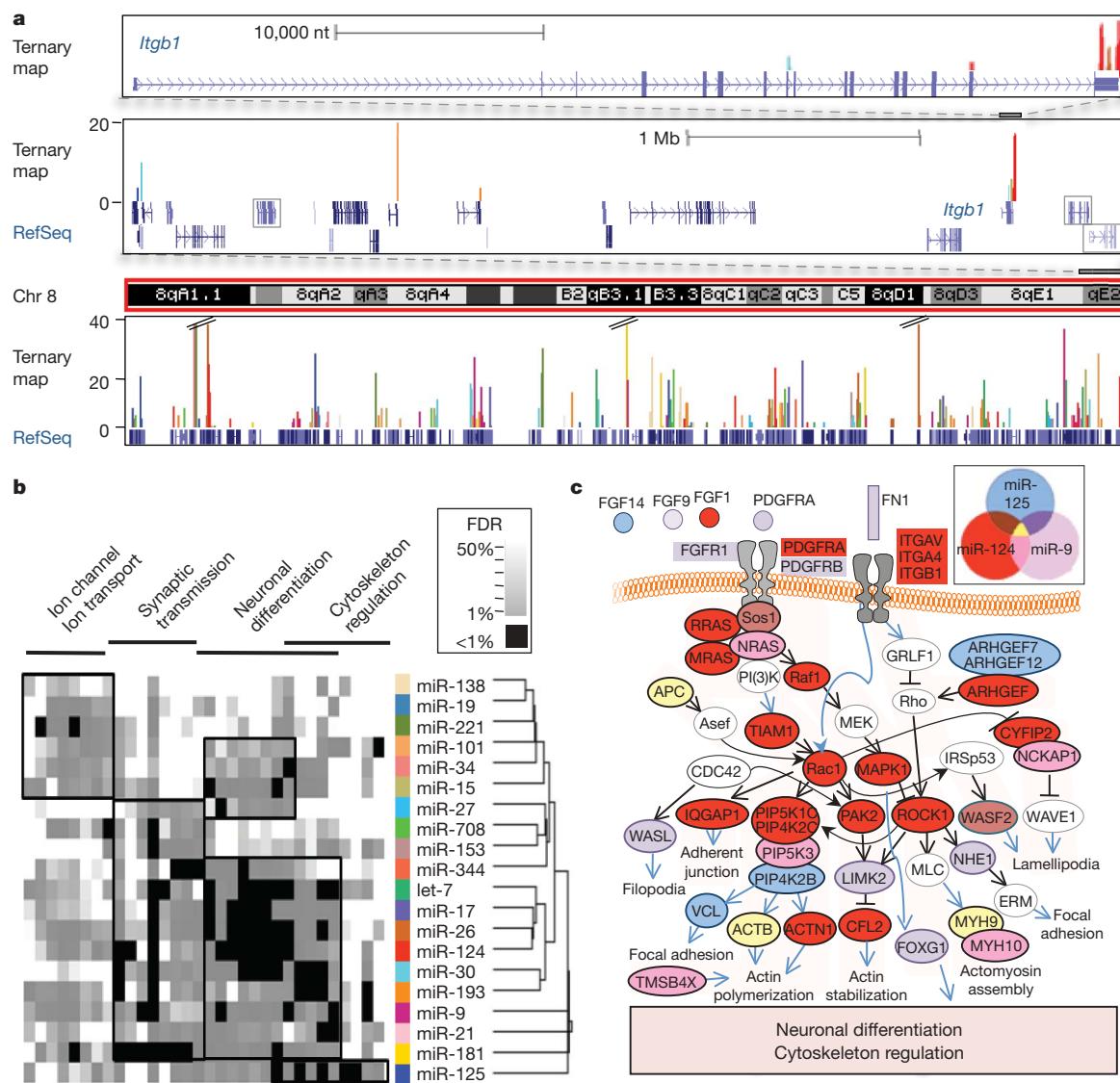


Figure 6 | Ago-miRNA ternary maps. **a**, Genome-wide views of Ago-miRNA ternary maps for the top 20 miRNAs from Ago HITS-CLIP (colours represent individual miRNA targets as indicated in **b**) are shown for the *Itgb1* gene (top panel), the local gene region (middle panel; all transcripts are expressed in P13 brain except those outlined in grey boxes), showing tags in neighbouring 3' UTRs, and for all of chromosome 8 (bottom panel). **b**, Heat map derived from gene ontology (GO) analysis of transcripts identified as targets of each of the top 20 miRNAs. The tree shows the

hierarchical clustering of miRNAs based on GO (Supplementary Methods). Significant clusters are outlined with black boxes (see Supplementary Fig. 12). **c**, Ago HITS-CLIP targets are shown for the most significant pathways (neuronal differentiation/cytoskeleton regulation; on the basis of false discovery rate (FDR), panel **b**) for miR-124, miR-9 and miR-125 in mouse brain. Actin cytoskeleton pathways are shown based on the KEGG database (<http://www.genome.jp/kegg/>).

predictions made using GO analysis of TargetScan predictions (Supplementary Fig. 15) demonstrates that the false discovery rate and ‘quality’ of the protein network deteriorate substantially when the Ago–mRNA map is not used. Target predictions from the Ago HITS-CLIP map suggest that diverse neuronal functions are regulated by different sets of miRNAs (Fig. 6b). The largest set of miRNA-associated functions, ‘neuronal differentiation’, illustrates interwoven but distinct pathways predicted to be regulated by three miRNAs expressed in neurons (Fig. 6c). The Ago–RNA ternary map corresponds remarkably well with the current view of miR-124, miR-125 and miR-9 biology, including actions to promote neurite outgrowth and differentiation by inhibiting Ago–miR-124 targets (Fig. 6c; discussed in Supplementary Fig. 15).

Discussion

Ago–miRNA–mRNA ternary maps identify functionally relevant miRNA binding sites in living tissues, and were developed in the context of several recent studies. Crystallographic structures of Ago–miRNA–mRNA ternary complexes³¹ demonstrated close contacts between all three molecules, consistent with the ability of CLIP, which requires close protein–RNA contacts^{28,29}, to detect both Ago–miRNA and Ago–mRNA interactions. The development of HITS-CLIP set the stage for generating and analysing genome-wide RNA–protein maps in the brain²⁶ and cultured cells³⁰. High throughput experiments and bioinformatic analysis together generated genome-wide predictions of miRNA seed sequences, particularly of miR-124. These studies demonstrated that miR-124 simultaneously represses hundreds of transcripts^{7,11,21,42–44}, and provided a genome-wide ‘gold standard’ with which to compare Ago HITS-CLIP data. This allowed estimates of specificity, false-positive and false-negative rates ($\sim 93\%$, $\sim 13\text{--}27\%$ and $\sim 15\text{--}25\%$, respectively; Supplementary Methods) that, although limited to one Ago–miRNA–mRNA data set, indicate that experimental Ago HITS-CLIP data outperforms bioinformatic predictions alone (Fig. 4 and Supplementary Fig. 11).

Although we used seed-driven approaches to validate targets, not all Ago binding need be constrained by these rules. Twenty-seven per cent of Ago–mRNA clusters have no predicted seed matches among the top 20 Ago–miRNA families. Such orphan clusters might bind other miRNAs, or miRNAs that follow other rules of binding, such as wobble or bulge nucleotides^{40,45,46} (S.W.C. and R.B.D., unpublished observations). Orphan clusters also provide another means of estimating the false-negative rate ($\sim 15\%$; Supplementary Methods), which compares favourably with previous studies in which false-negative rates were between 50% and 70%^{11,18,20}.

Ago HITS-CLIP resolves some obstacles that have arisen in efforts to understand miRNA action. It has been difficult to discriminate direct from indirect actions of miRNAs, and to extrapolate miRNA over-expression studies in tissue culture to organismal miRNA action. Target RNAs have previously been identified by immunoprecipitation, microarray analysis^{21,44} and reporter validation assays, with the concern that low stringency immunoprecipitation of non-crosslinked RNA–protein complexes⁴⁷, including Ago–miRNAs⁴⁸, may purify indirect targets. This has spurred interest in efforts to explore miRNA-target identification by covalently crosslinking, using formaldehyde or 4-thiouridine-modified RNA in culture to identify transcripts complexed with Ago, miRNAs and additional proteins^{48,49}. HITS-CLIP offers a clear means of identifying direct Ago targets and identifying specific interaction sites, which in turn offers the possibility of specifically targeting miRNA activity.

Ago HITS-CLIP complements bioinformatic approaches to miRNA target identification by restricting the sequence space to be analysed to the $\sim 45\text{--}60$ -nucleotide Ago footprint. For highly conserved 3' UTRs, such as those of the RNABPs Ptbp2, Nova1 and Fmr1, many miRNA sites are predicted using algorithms that rely on sequence conservation, but each has only one Ago–mRNA CLIP cluster (Supplementary Fig. 10). In fact, miRNA selectivity is very high such that on average, transcripts have between one to three

major Ago-binding sites in a single tissue (Fig. 3 and Supplementary Figs 10 and 14). Ago–mRNA binding sites themselves have no apparent sequence preference (data not shown), suggesting that accessibility may rely on additional RNABPs. Such a mechanism, which may be assessed by overlaying HITS-CLIP maps of different RNABPs²⁶, could provide a means of dynamically regulating miRNA binding and regulation¹.

By simultaneously generating binding maps for multiple miRNAs, Ago HITS-CLIP offers a new approach to understanding combinatorial control of target RNA expression. At the same time, analysis of a single miRNA, miR-124, demonstrated that its expression not only induced Ago to bind miR-124 sites, but reduced or precluded Ago binding to sites occupied in untransfected cells (Fig. 5), perhaps reflecting competition between a limited capacity for miRNA binding on a given 3' UTR⁵⁰. Such Ago occlusion has important mechanistic, experimental and clinical implications, where studies manipulating miRNA levels are envisioned.

METHODS SUMMARY

Ago HITS-CLIP was performed in biological replicates as described^{26,27} (using monoclonal antibody 2A8 or 7G1-1* as described in Supplementary Methods). High-throughput sequencing was performed with an Illumina Genome Analyser.

Affymetrix exon arrays (MoEx 1.0 ST) were used to measure transcript abundance in P13 mouse brain and data were analysed with Affymetrix Power Tools. Bioinformatics analysis used the UCSC genome browser, miRBase, BioPython, Scipy and GoMiner, as described in Supplementary Methods. Additional data access can be found at our project website <http://ago.rockefeller.edu/>.

Received 5 April; accepted 29 May 2009.

Published online 17 June 2009.

1. Filipowicz, W., Bhattacharyya, S. N. & Sonenberg, N. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nature Rev. Genet.* **9**, 102–114 (2008).
2. Sharp, P. A. The centrality of RNA. *Cell* **136**, 577–580 (2009).
3. Ambros, V. The functions of animal microRNAs. *Nature* **431**, 350–355 (2004).
4. Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281–297 (2004).
5. He, L. & Hannon, G. J. MicroRNAs: small RNAs with a big role in gene regulation. *Nature Rev. Genet.* **5**, 522–531 (2004).
6. Berezikov, E. et al. Diversity of microRNAs in human and chimpanzee brain. *Nature Genet.* **38**, 1375–1377 (2006).
7. Lim, L. P. et al. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* **433**, 769–773 (2005).
8. Krek, A. et al. Combinatorial microRNA target predictions. *Nature Genet.* **37**, 495–500 (2005).
9. Lewis, B. P., Burge, C. B. & Bartel, D. P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**, 15–20 (2005).
10. Rajewsky, N. microRNA target predictions in animals. *Nature Genet.* **38** (Suppl.), S8–S13 (2006).
11. Baek, D. et al. The impact of microRNAs on protein output. *Nature* **455**, 64–71 (2008).
12. Bentwich, I. Prediction and validation of microRNAs and their targets. *FEBS Lett.* **579**, 5904–5910 (2005).
13. Ventura, A. et al. Targeted deletion reveals essential and overlapping functions of the miR-17–92 family of miRNA clusters. *Cell* **132**, 875–886 (2008).
14. Ebert, M. S., Neilson, J. R. & Sharp, P. A. MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells. *Nature Methods* **4**, 721–726 (2007).
15. Johnston, R. J. & Hobert, O. A microRNA controlling left/right neuronal asymmetry in *Ceaeorhabditis elegans*. *Nature* **426**, 845–849 (2003).
16. Xiao, C. & Rajewsky, K. MicroRNA control in the immune system: basic principles. *Cell* **136**, 26–36 (2009).
17. van Rooij, E., Liu, N. & Olson, E. N. MicroRNAs flex their muscles. *Trends Genet.* **24**, 159–166 (2008).
18. Selbach, M. et al. Widespread changes in protein synthesis induced by microRNAs. *Nature* **455**, 58–63 (2008).
19. Beitzinger, M., Peters, L., Zhu, J. Y., Kremmer, E. & Meister, G. Identification of human microRNA targets from isolated argonaute protein complexes. *RNA Biol.* **4**, 76–84 (2007).
20. Easow, G., Teleman, A. A. & Cohen, S. M. Isolation of microRNA targets by miRNP immunopurification. *RNA* **13**, 1198–1204 (2007).
21. Hendrickson, D. G., Hogan, D. J., Herschlag, D., Ferrell, J. E. & Brown, P. O. Systematic identification of mRNAs recruited to argonaute 2 by specific microRNAs and corresponding changes in transcript abundance. *PLoS ONE* **3**, e2126 (2008).

22. Zhang, L. et al. Systematic identification of *C. elegans* miRISC proteins, miRNAs, and mRNA targets by their interactions with GW182 proteins AIN-1 and AIN-2. *Mol. Cell* **28**, 598–613 (2007).
23. Hammell, M. et al. mirWIP: microRNA target prediction based on microRNA-containing ribonucleoprotein-enriched transcripts. *Nature Methods* **5**, 813–819 (2008).
24. van Rooij, E., Marshall, W. S. & Olson, E. N. Toward microRNA-based therapeutics for heart disease: the sense in antisense. *Circ. Res.* **103**, 919–928 (2008).
25. Wheeler, T. M., Lueck, J. D., Swanson, M. S., Dirksen, R. T. & Thornton, C. A. Correction of CIC-1 splicing eliminates chloride channelopathy and myotonia in mouse models of myotonic dystrophy. *J. Clin. Invest.* **117**, 3952–3957 (2007).
26. Licatalosi, D. D. et al. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* **456**, 464–469 (2008).
27. Ule, J. et al. CLIP identifies Nova-regulated RNA networks in the brain. *Science* **302**, 1212–1215 (2003).
28. Ule, J., Jensen, K., Mele, A. & Darnell, R. B. CLIP: a method for identifying protein-RNA interaction sites in living cells. *Methods* **37**, 376–386 (2005).
29. Jensen, K. B. & Darnell, R. B. CLIP: crosslinking and immunoprecipitation of *in vivo* RNA targets of RNA-binding proteins. *Methods Mol. Biol.* **488**, 85–98 (2008).
30. Yeo, G. W. et al. An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. *Nature Struct. Mol. Biol.* **16**, 130–137 (2009).
31. Wang, Y. et al. Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. *Nature* **456**, 921–926 (2008).
32. Landgraf, P. et al. A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* **129**, 1401–1414 (2007).
33. Lu, J. et al. MicroRNA expression profiles classify human cancers. *Nature* **435**, 834–838 (2005).
34. Grimson, A. et al. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cell* **27**, 91–105 (2007).
35. Wang, E. T. et al. Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470–476 (2008).
36. Duursma, A. M., Kedde, M., Schrier, M., le Sage, C. & Agami, R. miR-148 targets human DNMT3b protein coding region. *RNA* **14**, 872–877 (2008).
37. Forman, J. J., Legesse-Miller, A. & Collier, H. A. A search for conserved sequences in coding regions reveals that the let-7 microRNA targets Dicer within its coding sequence. *Proc. Natl Acad. Sci. USA* **105**, 14879–14884 (2008).
38. Lal, A. et al. p16^{INK4a} translation suppressed by miR-24. *PLoS ONE* **3**, e1864 (2008).
39. Shen, W. F., Hu, Y. L., Uttarwar, L., Passegue, E. & Largman, C. MicroRNA-126 regulates HOXA9 by binding to the homeobox. *Mol. Cell. Biol.* **28**, 4609–4619 (2008).
40. Tay, Y., Zhang, J., Thomson, A. M., Lim, B. & Rigoutsos, I. MicroRNAs to *Nanog*, *Oct4* and *Sox2* coding regions modulate embryonic stem cell differentiation. *Nature* **455**, 1124–1128 (2008).
41. Yu, J. Y., Chung, K. H., Deo, M., Thompson, R. C. & Turner, D. L. MicroRNA miR-124 regulates neurite outgrowth during neuronal differentiation. *Exp. Cell Res.* **314**, 2618–2633 (2008).
42. Makeyev, E. V., Zhang, J., Carrasco, M. A. & Maniatis, T. The MicroRNA miR-124 promotes neuronal differentiation by triggering brain-specific alternative pre-mRNA splicing. *Mol. Cell* **27**, 435–448 (2007).
43. Wang, X. & Wang, X. Systematic identification of microRNA functions by combining target prediction and expression profiling. *Nucleic Acids Res.* **34**, 1646–1652 (2006).
44. Karginov, F. V. et al. A biochemical approach to identifying microRNA targets. *Proc. Natl Acad. Sci. USA* **104**, 19291–19296 (2007).
45. Didiano, D. & Hobert, O. Perfect seed pairing is not a generally reliable predictor for miRNA-target interactions. *Nature Struct. Mol. Biol.* **13**, 849–851 (2006).
46. Vella, M. C., Choi, E. Y., Lin, S. Y., Reinert, K. & Slack, F. J. The *C. elegans* microRNA let-7 binds to imperfect let-7 complementary sites from the lin-41 3'UTR. *Genes Dev.* **18**, 132–137 (2004).
47. Mili, S. & Steitz, J. A. Evidence for reassociation of RNA-binding proteins after cell lysis: implications for the interpretation of immunoprecipitation analyses. *RNA* **10**, 1692–1694 (2004).
48. Kirino, Y. & Mourelatos, Z. Site-specific crosslinking of human microRNPs to RNA targets. *RNA* **14**, 2254–2259 (2008).
49. Vasudevan, S., Tong, Y. & Steitz, J. A. Switching from repression to activation: microRNAs can up-regulate translation. *Science* **318**, 1931–1934 (2007).
50. Khan, A. A. et al. Transfection of small RNAs globally perturbs gene regulation by endogenous microRNAs. *Nature Biotechnol.* doi:10.1038/nbt.1543 (24 May 2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgments We thank members of the Darnell laboratory for discussion, J. Fak for help with exon arrays, C. Zhang for bioinformatic discussions, and G. Dunn, D. Licatalosi, C. Marney, M. Frias, T. Eom, J. Darnell, M. Yano and C. Zhang for critical review of the manuscript. We also thank Z. Mourelatos for supplying the ZA8 antibody and communicating unpublished results; G. Hannon for discussions; and S. Dewell for help with high-throughput sequencing. This work was supported in part by grants from the NIH (R.B.D.), the Cornell/Rockefeller/Sloan-Kettering Tri-Institutional Program in Computational Biology and Medicine (S.W.C.) and MD-PhD Program (J.B.Z.). R.B.D. is an Investigator of the Howard Hughes Medical Institute.

Author Contributions S.W.C. and R.B.D conceived, designed and supervised the experiments, analysed the data and wrote the paper. J.B.Z. did the initial experiments with the 7G1-1* antibody. A.M. helped with all HITS-CLIP experiments.

Author Information The microarray data have been deposited in the GEO database under accession number GSE16338. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to R.B.D. (darnelr@rockefeller.edu).

LETTERS

Liquid water on Enceladus from observations of ammonia and ^{40}Ar in the plume

J. H. Waite Jr¹, W. S. Lewis¹, B. A. Magee¹, J. I. Lunine², W. B. McKinnon³, C. R. Glein⁴, O. Mousis^{2,5}, D. T. Young¹, T. Brockwell¹, J. Westlake¹, M.-J. Nguyen¹, B. D. Teolis¹, H. B. Niemann⁶, R. L. McNutt Jr⁷, M. Perry⁷ & W.-H. Ip⁸

Jets of water ice from surface fractures near the south pole¹ of Saturn's icy moon Enceladus produce a plume of gas and particles^{2–5}. The source of the jets may be a liquid water region under the ice shell—as suggested most recently by the discovery of salts in E-ring particles derived from the plume⁶—or warm ice that is heated, causing dissociation of clathrate hydrates⁷. Here we report that ammonia is present in the plume, along with various organic compounds, deuterium and, very probably, ^{40}Ar . The presence of ammonia provides strong evidence for the existence of at least some liquid water, given that temperatures in excess of 180 K have been measured near the fractures from which the jets emanate⁸. We conclude, from the overall composition of the material, that the plume derives from both a liquid reservoir (or from ice that in recent geological time has been in contact with such a reservoir) as well as from degassing, volatile-charged ice.

Enceladus' plume has been sampled with the Cassini Saturn Orbiter's Ion and Neutral Mass Spectrometer (INMS)⁹ during five fly-bys of the moon and found to be composed predominantly of H_2O , with CO_2 as the second most abundant species³. The encounter on 9 October 2008 yielded data with the highest signal-to-noise ratio obtained so far, enabling the identification of trace species, including complex organics such as benzene, that were not distinguishable in INMS data from the other fly-bys (Table 1 and Fig. 1). (Spectra and mixing ratios from the other encounters are presented in Supplementary Information.) Of particular importance for constraining models of Enceladus' interior and origin is the definitive detection of ammonia (NH_3), the probable detection of radiogenic argon, and the measurement of the deuterium-to-hydrogen ratio in H_2O .

Ammonia is present in the plume with a mixing ratio of 0.8%. (Previously we reported only an upper limit for NH_3 of <0.5%; ref. 3.) The NH_3 was probably present in the planetesimals from which Enceladus formed, and we note that its detection at Enceladus circumstantially supports the postulated presence of NH_3 in Titan^{10–12}. Ammonia's presence in the plume, along with the detection of Na and K salts in E-ring ice particles⁶, implies that the interior of Enceladus may contain some amount of liquid water. Ammonia (together with methanol and salts) acts as an antifreeze that permits the existence of liquid water down to temperatures as low as 176 K. Preservation of even a residual oceanic layer during cycles of global cooling¹³ would maintain the conditions necessary for renewed episodes of tidal heating and geological activity¹⁴.

As is the case at Titan^{10,11}, the low upper limit determined for primordial argon in the plume (see below) suggests that NH_3 rather than N_2 -bearing clathrate was the major carrier of nitrogen in the

material from which Enceladus formed. Thermal decomposition of NH_3 in Enceladus' interior would produce N_2 (ref. 15), which, if present in the plume, would contribute to the signal at mass 28 (that is, 28 Da) in the mass spectrum. The interpretation of this mass 28 peak is however uncertain. The signal can be attributed, with equal plausibility, to either N_2 or C_2H_4 (or some combination thereof) as well as to CO from the hypervelocity impact dissociation of CO_2 plus a possible admixture of native CO (see Supplementary Information). The presence of C_2H_4 in the plume would be puzzling, as ethylene has

Table 1 | INMS determination of plume composition on 9 October 2008

Species	Volume mixing ratio
H_2O	0.90 ± 0.01
CO_2	0.053 ± 0.001
CO	[0.044]
H_2	[0.39]
H_2CO	$(3.1 \pm 1) \times 10^{-3}$
CH_3OH	$(1.5 \pm 0.6) \times 10^{-4}$
$\text{C}_2\text{H}_4\text{O}$	$< 7.0 \times 10^{-4}$
$\text{C}_2\text{H}_6\text{O}$	$< 3.0 \times 10^{-4}$
H_2S	$(2.1 \pm 1) \times 10^{-5}$
^{40}Ar	$(3.1 \pm 0.3) \times 10^{-4}$
NH_3	$(8.2 \pm 0.2) \times 10^{-3}$
N_2	< 0.011
HCN^\dagger	$< 7.4 \times 10^{-3}$
CH_4	$(9.1 \pm 0.5) \times 10^{-3}$
C_2H_2	$(3.3 \pm 2) \times 10^{-3}$
C_2H_4	< 0.012
C_2H_6	$< 1.7 \times 10^{-3}$
C_3H_4	$< 1.1 \times 10^{-4}$
C_3H_6	$(1.4 \pm 0.3) \times 10^{-3}$
C_3H_8	$< 1.4 \times 10^{-3}$
C_4H_2	$(3.7 \pm 0.8) \times 10^{-5}$
C_4H_4	$(1.5 \pm 0.6) \times 10^{-5}$
C_4H_6	$(5.7 \pm 3) \times 10^{-5}$
C_4H_8	$(2.3 \pm 0.3) \times 10^{-4}$
C_4H_{10}	$< 7.2 \times 10^{-4}$
C_5H_6	$< 2.7 \times 10^{-6}$
C_5H_{12}	$< 6.2 \times 10^{-5}$
C_6H_6	$(8.1 \pm 1) \times 10^{-5}$

The mixing ratios shown here for CO and H_2 (values in square brackets) are included in the mixing ratios for CO_2 and H_2O given in the first two rows. Analysis of the data from all five encounters shows that the ratios of mass 44 (CO_2) to mass 28 and of mass 18 (H_2O) to mass 2 (H_2) decrease with increasing spacecraft velocity, suggesting that H_2 and CO are produced by the dissociation of H_2O and CO_2 through hypervelocity impact on, and reaction with, the walls of the INMS antechamber (see Supplementary Information). We estimate that 40–80% of the signal in mass channel 28 is due to CO produced in this way. A small contribution of CO from Enceladus is also possible, but cannot be distinguished from the dissociation product. The residual mass 28 signal is attributed to N_2 , C_2H_4 , or a combination of both, with either HCN or the ethene dissociation product C_2H_3 contributing to the signal in mass channel 27. The values given for these species are upper limits based on the two alternative scenarios ($\text{N}_2 + \text{HCN}$ versus C_2H_4). Neither scenario can be given preference over the other on the basis of the present INMS data set.

¹Southwest Research Institute, San Antonio, Texas 78228, USA. ²Lunar and Planetary Laboratory, University of Arizona, Tucson, Arizona 85721, USA. ³Department of Earth and Planetary Sciences, Washington University, St Louis, Missouri 63130, USA. ⁴School of Earth and Space Exploration, Arizona State University, Tempe, Arizona 85287, USA.

⁵Observatoire de Besançon, Institut UTINAM, CNRS-UMR 6213, BP 1615, 25010 Besançon Cedex, France. ⁶NASA Goddard Space Flight Center, Greenbelt, Maryland 20771, USA. ⁷The Johns Hopkins University Applied Physics Laboratory, Laurel, Maryland 20723, USA. ⁸Institutes of Astronomy and Space Science, National Central University, Chung Li 32054, Taiwan.

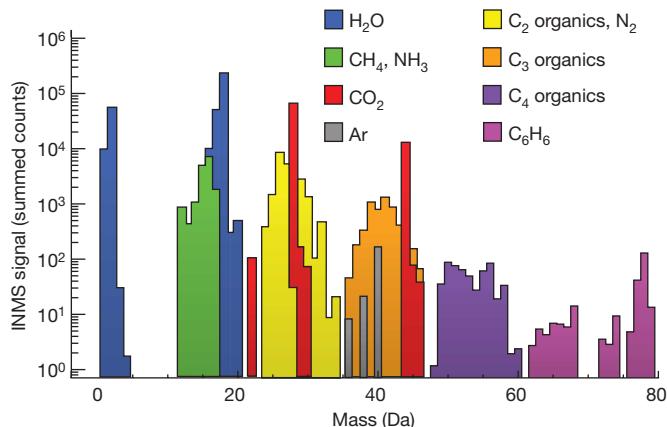


Figure 1 | Mass spectrum on 9 October 2008. The first INMS measurements of the plume's composition were made in the plume's outer edge on 15 July 2005 (the E2 encounter), and revealed that the plume consists predominantly of water vapour, with small amounts of carbon dioxide, methane, a species of mass 28, and trace quantities of acetylene and propane. During the 9 October 2008 (E5) encounter, favourable spacecraft pointing and the high spacecraft velocity relative to Enceladus significantly increased the neutral gas density in the closed source antechamber. This density enhancement, together with the proximity of the measurements to the centre of the plume, enabled a considerably improved signal-to-noise ratio compared with that obtained during the other four encounters. (The *y* axis shows counts summed over the period of plume influence.) With their higher signal-to-noise ratio, the E5 measurements provide the first clear evidence of the presence of higher-order (C_{4+}) hydrocarbons in the plume, including benzene, and allow the tentative identification of individual species within each hydrocarbon group and of other organics, such as methanol and formaldehyde (Table 1). As described in the Supplementary Information, interpretation of the INMS spectra requires careful deconvolution of a complex pattern of mass peaks representing both parent species and dissociative ionization products. Fractionation by energetic impact as well as interaction of the gas with the titanium antechamber walls must also be taken into account. Owing to the low mass resolution of INMS, overlapping mass peaks cannot in a number of cases be uniquely separated, resulting in some ambiguity in the interpretation of the data.

not been detected in comets¹⁶ and is thus unlikely to have been present in large quantities in the icy planetesimals from which Enceladus formed. The presence of N_2 can be more readily explained as resulting from the hydrothermal processing of NH_3 (refs 15, 17). Complicating this scenario is the presence of HCN, which is required to fit the spectrum at mass 27. HCN is rapidly hydrolysed to formic acid and NH_3 in warm water¹⁸. Hence its preservation on Enceladus and presence in the plume may be incompatible with the hydrothermal production of N_2 . On the other hand, the coexistence of N_2 and HCN may mean that the plume consists of a mixture of materials whose sources experienced different degrees of aqueous processing, including primordial material trapped in ice that has not been in contact with liquid water.

The optimum fit to the spectrum in the mass range 36–43 Da leaves a significant residual signal at mass 40 (Fig. 2), which we attribute to the presence of radiogenic ^{40}Ar at a volume mixing ratio of 3×10^{-4} . (^{36}Ar and ^{38}Ar —primordial non-radiogenic argon—may also be present, but at a combined abundance of $< 10^{-5}$.) This amount of ^{40}Ar is greater by three orders of magnitude than would be expected for a chondritic abundance of potassium in Enceladus' rock fraction¹⁹, thus requiring both an efficient mechanism for the escape of ^{40}Ar from the rock component and a mechanism for concentrating it. In an undifferentiated rock–ice body, ^{40}Ar could escape through diffusion from small particles over several Gyr; however, solid-state diffusion is inefficient over larger length scales, and this scenario does not provide a means of concentrating ^{40}Ar within the ice. In contrast, water–rock interactions early in Enceladus' history would naturally facilitate ^{40}Ar escape by leaching both ^{40}Ar and ^{40}K from a rock core

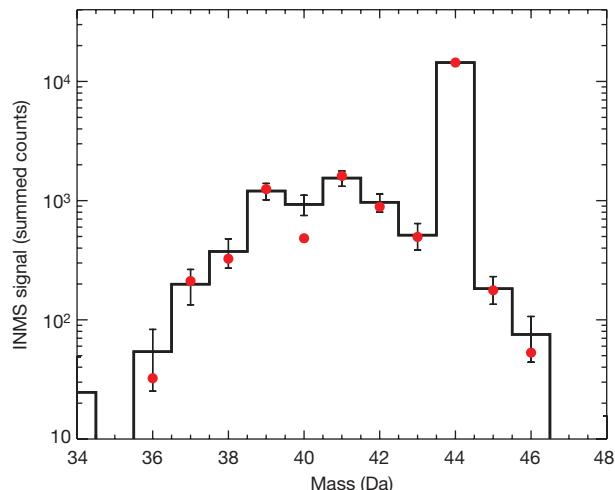


Figure 2 | Histogram showing the INMS spectrum over the mass range 34–48 Da. The red dots show a fit consisting of a mix of CO_2 , C_3H_4 , C_3H_6 and C_3H_8 . The residual signal at mass 40 that we attribute to ^{40}Ar is clearly evident. C_3H_4 has a strong main peak at 40 Da; however, the contribution of this species to the signal in mass channel 40 is constrained by its contributions to the signal in channels 36–39. Besides C_3H_8 , the species C_4H_{10} and C_2H_4O contribute significant signal to mass 43 while maintaining a good fit across the rest of the spectrum; only negligible changes occur in the fit if C_4H_{10} and/or C_2H_4O are substituted for C_3H_8 . The error bars indicate the 1σ error associated with the counting statistics of the summed signals.

into an oceanic layer²⁰. Moreover, as long as the volume of water in which the ^{40}Ar ends up was much less than the volume of the (presumably) overlying ice shell, the ocean water would become enriched in ^{40}Ar . ^{40}Ar -enriched water may thus be directly involved in the observed plume activity; alternatively, the plume may access degassing clathrate hydrate that formed from ^{40}Ar -enriched ocean water on geologically recent timescales and is therefore itself enriched.

At today's H_2O loss rate, as measured by the Cassini Ultraviolet Imaging Spectrograph (UVIS)⁴, the total reservoir of ^{40}Ar produced by radioactive decay over Enceladus' history would be exhausted in less than 10 Myr. This suggests that the observed amount of ^{40}Ar is not a steady-state value or that plume activity is intermittent, with a duty cycle of $< 1\%$ (or both). Such a duty cycle may be necessary to reconcile Enceladus' observed heat flow with the long-term average available from tidal dissipation²¹.

The D/H ratio— $2.9 (+1.5/-0.7) \times 10^{-4}$ —was determined from H_2 and HD, assumed to be products of H_2O dissociation by reaction with antechamber titanium and therefore representative of the plume H_2O inventory (see Supplementary Information). The D/H ratio is close to the cometary value of $\sim 3 \times 10^{-4}$ (ref. 16), nearly twice the terrestrial ocean water value (1.56×10^{-4})¹⁹, and more than ten times the value of the D/H ratio in the protosolar nebula ($2.1 \pm 0.4 \times 10^{-5}$)²². The comet-like ratio provides definitive evidence that Enceladus accreted from planetesimals formed in the solar nebula rather than in an initially dense, warm subnebula. In the latter case, D/H in H_2O would be lower because HDO would have isotopically exchanged with H_2 before the water vapour condensed in the subnebula²³. Thus the higher D/H ratio observed in the plume favours the hypothesis that Saturn's regular satellites formed in a low-surface-density circumplanetary disk generated during the last phase of the planet's growth and fed by gas and gas-coupled solids originating from the solar nebula^{23–25}.

In addition to the D/H ratio, Enceladus' plume displays other striking compositional similarities to cometary comae, although notable differences exist as well (Fig. 3), such as the relative deficiency of native CO (see Supplementary Information) and the ubiquity of N_2 and/or C_2H_4 in the plume (Table 1). The planetesimals from which Enceladus formed may not have contained large quantities of CO because they were thermally processed in Saturn's disk²⁵, or accreted

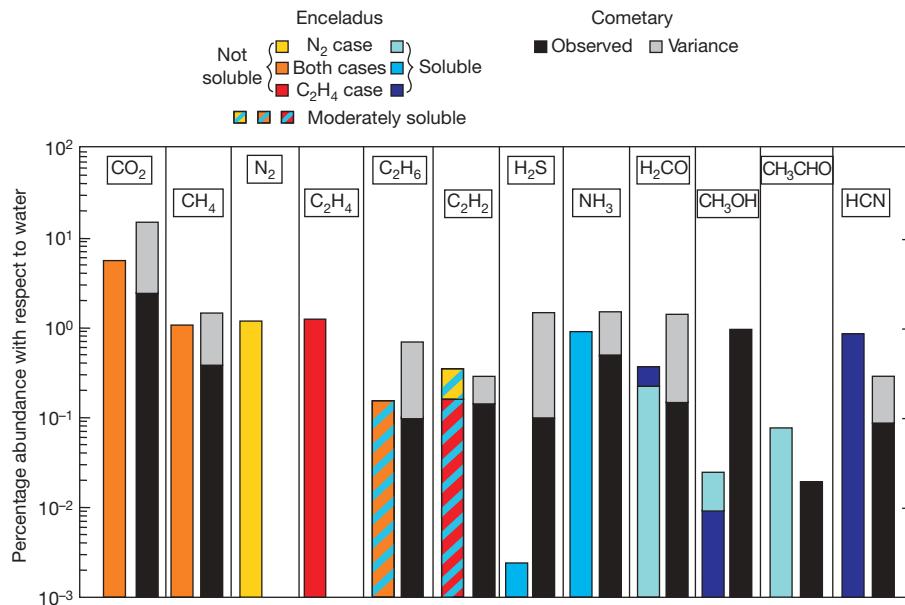


Figure 3 | Comparison of volatile abundances observed in Enceladus' plume with those seen in cometary comae. (Comet data from ref. 16.) The abundances of CO₂, CH₄, C₂H₂, NH₃ and H₂CO (relative to H₂O) in the plume resemble those in comets, suggesting that these abundances are typical of well-mixed outer Solar System material. The plume may contain proportions of CH₃CHO and HCN that are somewhat elevated compared to comets. In contrast, the plume H₂S and CH₃OH abundances are significantly lower than those in comets. The most conspicuous difference is

CO may have been hydrothermally processed²⁶. (UVIS did not detect CO in the plume, establishing a 2 σ upper limit of <3%; ref. 27.)

The composition of the plume revealed by the INMS data attests to the complex history and present state of Enceladus. Incomplete melting of Enceladus' ice shell and (past) contact between a hot rocky core and an oceanic layer may have resulted in a dynamic situation, so that we are measuring the chemical output of a disequilibrium, multiphase system rather than one in thermodynamic equilibrium. The measurements reported here, together with observations of sodium salts in E-ring particles⁶ and the gas velocities in the plumes²⁸, provide compelling evidence that a liquid water reservoir exists—or recently (on geological timescales) existed—in the interior.

Various alternative plume sources have been proposed, including a subsurface reservoir of boiling water¹ and the sublimation of warm ice or clathrates^{7,29}. The INMS data seem to require both liquid water and solid ice³⁰. On the one hand, the abundant presence of ⁴⁰Ar, and of NH₃ as well as perhaps N₂, argues for aqueous activity in Enceladus' interior. But on the other hand, the abundances of CO₂ and especially CH₄ relative to H₂O are higher than the aqueous solubilities of these compounds at conditions representative of Enceladus' interior (Fig. 3 and Supplementary Table 4). It is possible to imagine that a liquid reservoir is a multiphase system in which the liquid water phase coexists with condensed or gaseous phases of those constituents. Alternatively, ice above an aqueous zone, such as a clathrate hydrate⁷, could be contributing those species to the plume, as well as all or some of C₃H₈, C₄H₈, C₄H₁₀, HCN and N₂/C₂H₄ (Supplementary Table 4).

Received 18 March; accepted 22 May 2009.

1. Porco, C. C. et al. Cassini observes the active south pole of Enceladus. *Science* **311**, 1393–1401 (2006).
2. Dougherty, M. K. et al. Identification of a dynamic atmosphere at Enceladus with the Cassini Magnetometer. *Science* **311**, 1406–1409 (2006).
3. Waite, J. H. Jr et al. Cassini Ion and Neutral Mass Spectrometer: Enceladus plume composition and structure. *Science* **311**, 1419–1422 (2006).
4. Hansen, C. J. et al. Enceladus' water vapor plume. *Science* **311**, 1423–1425 (2006).
5. Spahn, F. et al. Cassini dust measurements at Enceladus and implications for the origin of the E ring. *Science* **311**, 1416–1418 (2006).
6. Postberg, F. et al. Sodium salts in E-ring ice grains from an ocean below the surface of Enceladus. *Nature* **459**, 1098–1101 (2009).
7. Kieffer, S. W. et al. A clathrate reservoir hypothesis for Enceladus' south polar plume. *Science* **314**, 1764–1766 (2006).
8. Spencer, J. et al. High spatial resolution observations of thermal emission from Enceladus' active south pole. *Eos* **89** (Fall Meet. Suppl.), abstr. P23B-1372 (2008).
9. Waite, J. H. Jr et al. The Cassini Ion and Neutral Mass Spectrometer (INMS) investigation. *Space Sci. Rev.* **114**, 113–231 (2004).
10. Owen, T. On the origin of Titan's atmosphere. *Planet. Space Sci.* **48**, 747–752 (2000).
11. Niemann, H. B. et al. The abundances of constituents of Titan's atmosphere from the GCMS instrument on the Huygens probe. *Nature* **438**, 779–784 (2005).
12. Lopes, R. M. C. et al. Cryovolcanic features on Titan's surface as revealed by the Cassini Titan Radar Mapper. *Icarus* **186**, 395–412 (2007).
13. Roberts, J. H. & Nimmo, F. Tidal heating and the long-term stability of a subsurface ocean on Enceladus. *Icarus* **194**, 675–689 (2007).
14. Tobie, G., Čadek, O. & Sotin, C. Solid tidal friction above a liquid water reservoir as the origin of the south pole hotspot on Enceladus. *Icarus* **196**, 642–652 (2008).
15. Matson, D. L., Castillo, J. C., Lunine, J. I. & Johnson, T. V. Enceladus' plume: compositional evidence for a hot interior. *Icarus* **187**, 569–573 (2007).
16. Bockelée-Morvan, D., Crovisier, J., Mumma, M. J. & Weaver, H. A. in *Comets II* (eds Festou, M. C., Keller, H. U. & Weaver, H. A.) 391–423 (Univ. Arizona Press, 2004).
17. Glein, C. R., Zolotov, M. Y. & Shock, E. L. The oxidation state of hydrothermal systems on early Enceladus. *Icarus* **197**, 157–163 (2008).
18. Miyakawa, S. et al. The cold origin of life: A. Implications based on the hydrolytic stabilities of hydrogen cyanide and formamide. *Orig. Life Evol. Biosph.* **32**, 195–208 (2002).
19. Lodders, K. & Fegley, B. Jr *A Planetary Scientist's Companion* (Oxford Univ. Press, 1998).
20. Zolotov, M. Y. An oceanic composition on early and today's Enceladus. *Geophys. Res. Lett.* **34**, L23203, doi:10.1029/2007GL031234 (2007).
21. Meyer, J. & Wisdom, J. Tidal heating in Enceladus. *Icarus* **188**, 235–239 (2007).
22. Geiss, J. & Gloeckler, G. Abundances of deuterium and helium in the protosolar cloud. *Space Sci. Rev.* **84**, 239–250 (1998).
23. Horner, J., Mousis, O., Alibert, Y., Lunine, J. I. & Blanc, M. Constraints on the formation of icy bodies in the Jovian system and beyond. *Planet. Space Sci.* **56**, 1585–1595 (2008).
24. Canup, R. M. & Ward, W. R. Formation of the Galilean satellites: conditions of accretion. *Astron. J.* **124**, 3404–3423 (2002).
25. Mousis, O. et al. Clathration of volatiles in the solar nebula and implications for the origin of Titan's atmosphere. *Astrophys. J.* **691**, 1780–1786 (2009).
26. Shock, E. L. & McKinnon, W. B. Hydrothermal processing of cometary volatiles: applications to Triton. *Icarus* **106**, 464–477 (1993).

the deficiency of native CO relative to that produced in the instrument (see Supplementary Information) and the ubiquity of N₂/C₂H₄ in the plume. These features are puzzling, because CO is generally the most abundant non-H₂O volatile in comets (observed range 0.4–20%), and neither N₂ nor C₂H₄ has been detected in comets. The abundances of CO₂, CH₄, C₂H₂ and C₂H₄ exceed the aqueous solubilities of those species, suggesting that an aqueous medium coexists with other forms of these species, presumably in the ice overlying the medium.

27. Hansen, C. J. et al. Water vapour jets inside the plume of gas leaving Enceladus. *Nature* **456**, 477–479 (2008).
28. Schmidt, J., Brillantov, N., Spahn, F. & Kempf, S. Slow dust in Enceladus' plume from condensation and wall collisions in tiger stripe fractures. *Nature* **451**, 685–688 (2008).
29. Nimmo, F., Spencer, J. R., Pappalardo, R. T. & Mullen, M. E. Shear heating as the origin of the plumes and heat flux on Enceladus. *Nature* **447**, 289–291 (2007).
30. Fortes, A. D. Metasomatic clathrate xenoliths as a possible source for the south polar plumes of Enceladus. *Icarus* **191**, 743–748 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Support by the Cassini Project through a subcontract with Southwest Research Institute is acknowledged.

Author Contributions J.H.W. organized and coordinated the plume composition study. J.H.W., C.R.G., W.S.L., J.I.L., W.B.M., O.M., M.-J.N. and J.W. developed the concept of the Letter and the interpretation of the INMS data. W.S.L., J.I.L., W.B.M., C.R.G. and O.M. worked jointly to write the manuscript. B.A.M. analysed the data and prepared the spectra. D.T.Y., T.B. and B.D.T. analysed the physical and chemical interactions of the inflowing material with the INMS antechamber and, together with B.A.M. and C.R.G., prepared the Supplementary Information. H.B.N. provided calibration support and information about instrument performance. R.L.M., M.P. and W.-H.I. provided supporting analysis of the neutral and ion environment surrounding Enceladus.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to W.S.L. (wlewis@swri.edu).

LETTERS

Asymmetric auroral intensities in the Earth's Northern and Southern hemispheres

K. M. Laundal¹ & N. Østgaard¹

It is commonly assumed that the aurora borealis (Northern Hemisphere) and aurora australis (Southern Hemisphere) are mirror images of each other because the charged particles causing the aurora follow the magnetic field lines connecting the two hemispheres. The particles are believed to be evenly distributed between the two hemispheres, from the source region in the equatorial plane of the magnetosphere. Although it has been shown that similar auroral features in the opposite hemispheres can be displaced tens of degree in longitude^{1,2} and that seasonal effects can cause differences in global intensity^{3,4}, the overall auroral patterns were still similar. Here we report observations that clearly contradict the common assumption about symmetric aurora: intense spots are seen at dawn in the Northern summer Hemisphere, and at dusk in the Southern winter Hemisphere. The asymmetry is interpreted in terms of inter-hemispheric currents related to seasons, which have been predicted^{5,6} but hitherto had not been seen.

On 12 May 2001, the Wideband Imaging Camera (WIC)⁷ on the Imager for Magnetopause-to-Aurora Global Exploration (IMAGE) satellite and the Polar Visible Imaging System (VIS) Earth camera⁸ recorded completely asymmetric auroral intensity distributions in the two hemispheres, with higher intensity in the northern dawn, and in the southern dusk (Fig. 1). A spatiotemporal analysis (Fig. 2) reveals that the dusk aurora in the Southern Hemisphere was persistent, lasting until 22:30 Universal Time (UT) and that the northern dawn spot was a transient feature lasting about 10 min and occurring twice, at 21:43 and at 22:10 UT.

The charged particles causing the spots at dawn and dusk originate from different regions in the magnetosphere, as sketched in the inset in Fig. 1b. Although we do not have the data to address this properly, the particles creating the two spots are probably accelerated by parallel electric fields and/or Alfvén waves. The observed asymmetric intensity distribution may imply differences in acceleration in the two hemispheres.

It is well-known that the magnetic field strength in the two hemispheres is not the same and this may generate differences in auroral intensity between the hemispheres⁹. However, the differences in the magnetic field strength in the Earth's ionosphere where the spots are observed are very small (<10%) and are not likely to explain the large asymmetries we observe.

The difference in ionospheric conductivity is expected to give rise to different auroral intensities in the two hemispheres. If a generator in the magnetosphere drives a current through the ionosphere, low ionospheric conductivity needs to be compensated for by a larger electric field, which accelerates the precipitating particles. An upward current is associated with precipitating electrons. This effect, which favours the dark hemisphere where the conductivity is low, has been

observed statistically^{3,4}. In this case (Fig. 1b), the Southern Hemisphere was in darkness, and the Northern Hemisphere was sunlit. This implies a lower conductivity in the Southern Hemisphere, and we would expect to see more intense aurora in the south than in the north, which is true for the dusk spot but not for the dawn spot. The asymmetry we observe therefore cannot be fully explained by this mechanism.

We know from previous studies that the orientation of the interplanetary magnetic field can have a strong impact on auroral asymmetries^{1,2,10,11}, in which the B_y component has been found to be the most important, just as it is in the large-scale convection patterns¹². The interplanetary magnetic field conditions between 21:20–22:30 UT were dominated by a large positive B_x component, as sketched in Fig. 1b. Theoretical considerations¹³ suggest that a dominant positive interplanetary magnetic field B_x component implies that the solar wind dynamo is more efficient and drives stronger high latitude (region 1) currents in the Southern Hemisphere. This idea is consistent with the persistent dusk spot but cannot explain the transient feature in the north.

Theoretical studies^{5,6} predict that conductivity differences will lead to inter-hemispheric, field-aligned currents. However, these currents have never been observed. The theory predicts a pair of inter-hemispheric currents going up (down) at dusk and down (up) at dawn in the southern dark (northern sunlit) hemisphere. Furthermore, as these currents are related to conductivity gradients they should be observed in the vicinity of the terminator, which is consistent with our observations. The persistent character of the asymmetric southern dusk spot is enhanced by the more efficient solar wind dynamo in the Southern Hemisphere for large positive B_x . The spot at dawn coincides with a significant increase of tail reconnection in this region, which can explain its transient character. This also suggests that both Alfvén waves and plasma flows and shears might be involved in creating the bright aurora. We therefore think that the asymmetry we observe is a confirmation of these predicted currents that to our knowledge has not been observed in both hemispheres before.

All our observational knowledge about interhemispheric differences in the aurora comes from a few ground-based observations with narrow field of view¹⁴, statistical studies from one hemisphere^{3,4}, and some recent studies based on simultaneous imaging from space^{1,2,10,11}. The latter has been possible only twice in history, first with Viking and Dynamic Explorer 1 in the mid-1980s, and then with Polar and IMAGE. Because neither of these satellites were operated to provide conjugate images, the coverage is sporadic, and thus we do not know how common direct observations of interhemispheric currents could be. However, the most intense interhemispherical current systems should be observable from the ground and not only statistically¹⁵.

¹Department of Physics and Technology, Allegt 55, University of Bergen, N-5007 Bergen, Norway.

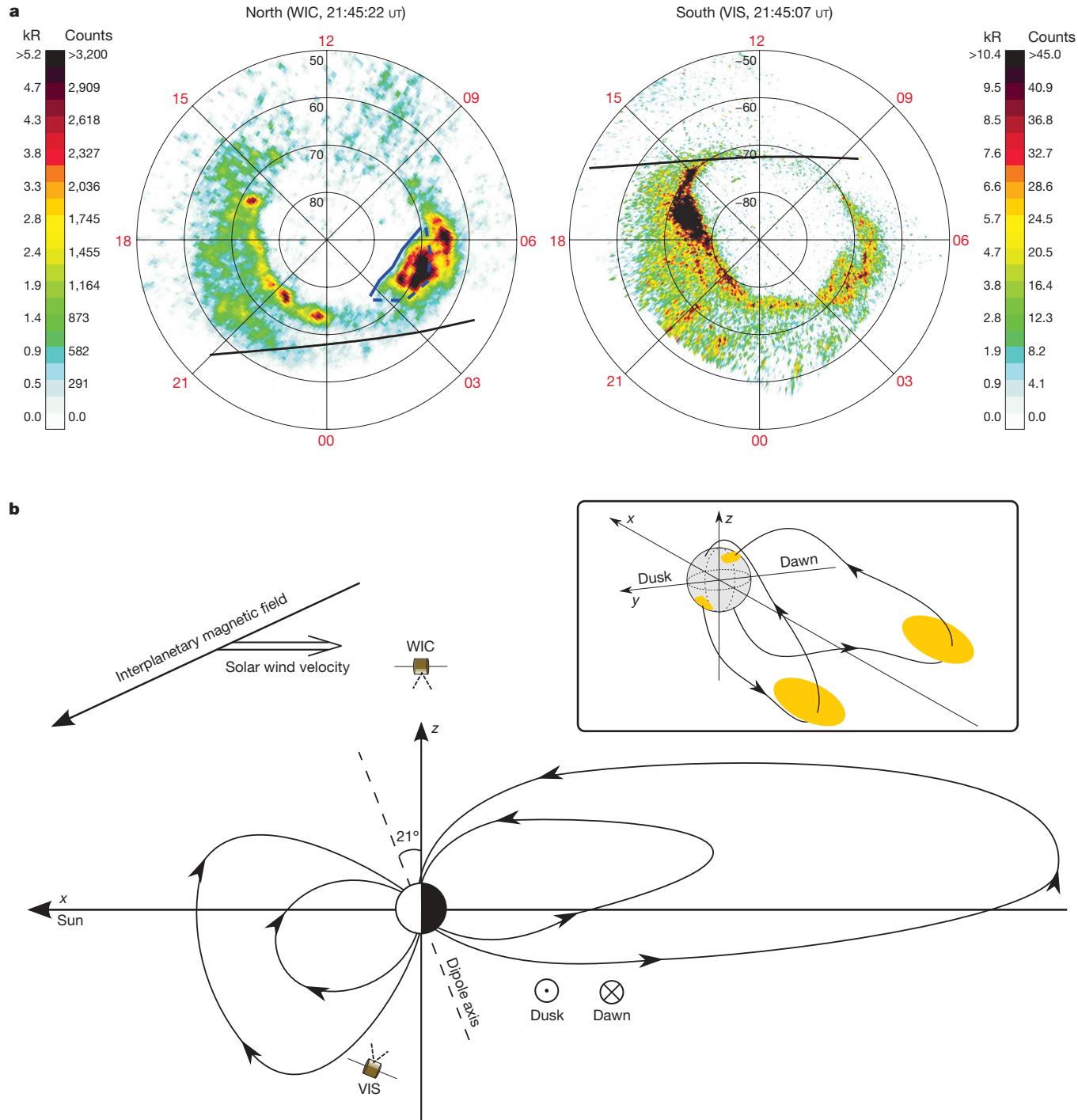


Figure 1 | Simultaneous ultraviolet images of the aurora in both hemispheres. **a**, Simultaneous images of the aurora at 21:45 UT 12 May 2001, showing completely different intensity distributions. WIC, image of the Northern Hemisphere from the IMAGE satellite's WIC camera; VIS, image of the Southern Hemisphere from the Polar VIS Earth Camera. The colour bars show intensity in both counts and kilo-Rayleigh (kR)^{16,17}. The exposure times for WIC and VIS are 10 s and 32.5 s, and the UT at the middle of the exposure is shown above the images; the cadences for WIC and VIS are 2 min and 1 min. The images are shown in magnetic apex coordinates looking down at the magnetic North Pole (and through the Earth for the southern image), with the direction of the Sun up, dawn to right, and dusk to the left. Sunlight-induced dayglow emissions have been subtracted from the images. The black curves indicate the sunlight terminator. The blue lines show the location of the poleward boundary of the aurora at 21:37 (dashed)

and 21:45 (solid) UT. This means that in less than 10 min the boundary, which is also the boundary between open and closed magnetic field lines, moved about 8° magnetic latitude towards the pole, signifying a local increase of tail reconnection. **b**, A conceptual presentation showing the seasonal conditions and geometry of the magnetosphere, and the orientation of the interplanetary magnetic field, measured by ACE ($|By|$ was less than 1 nT). The interplanetary magnetic field moves with the solar wind velocity. The Earth's magnetic dipole axis (shown with a dashed line) was tilted at an angle of 21° towards the Sun at 21:45 UT. The coordinate system in this sketch is Geocentric Solar Magnetospheric, with the x axis pointing towards the Sun, the y axis perpendicular to the Earth's magnetic dipole axis (and to the x axis) and the z axis fulfilling the system. The inset illustrates that the spots at northern dawn and southern dusk originate from completely different regions in the magnetosphere.

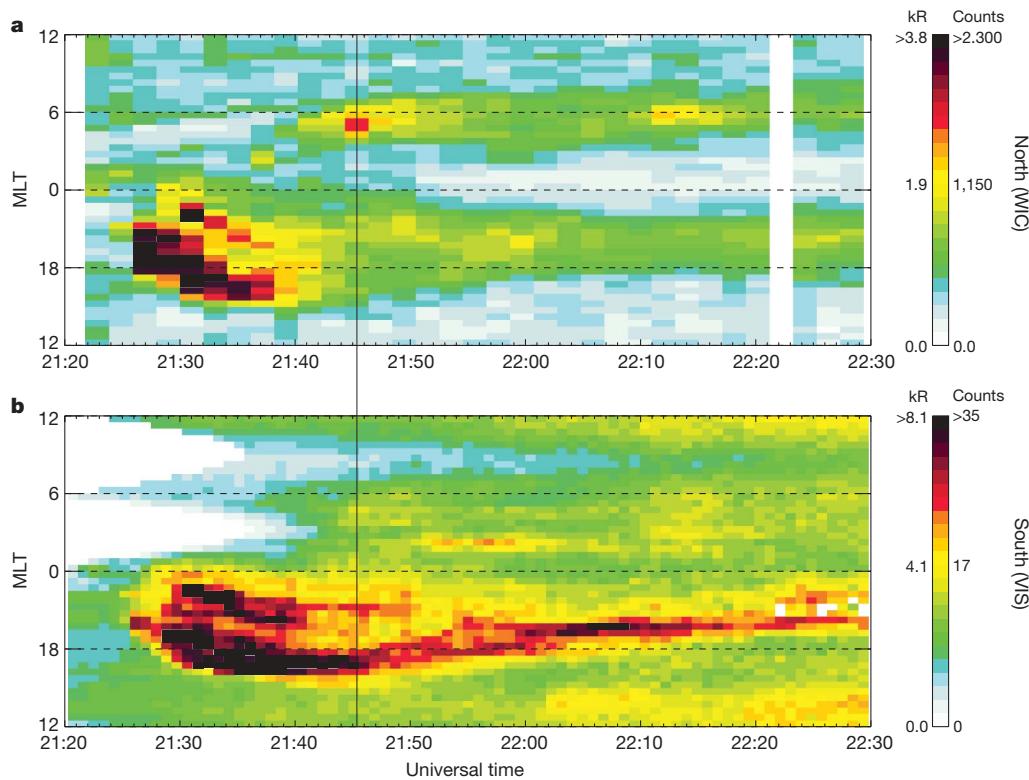


Figure 2 | Spatio-temporal distribution of auroral intensity in both hemispheres. The intensity (integrated between 63° and 85° magnetic latitude) is shown as a function of UT (*x* axis) and Magnetic Local Time (MLT, *y* axis), in the Northern (**a**) and Southern (**b**) hemispheres. Panel **b** reveals that the spot at southern dusk coincides with the western edge of an auroral substorm, seen clearly in both hemispheres between 21:25 and 21:40 UT. The count rate in both cameras depends on the energy flux of precipitating particles. However, the cameras have different sensitivity and pass-band, giving different counts and kR in the two hemispheres^{16,17}. The VIS camera (124–149 nm) is most sensitive to the atomic oxygen (O I) line at 130.4 nm (ref. 16), while the WIC camera (140–190 nm) is sensitive to molecular nitrogen (N_2) emissions in the Lyman–Birge–Hopfield (LBH) band and a few emission lines of atomic nitrogen^{7,17}. An important difference

in energy flux dependence between the two cameras is that emissions in the LBH band are more reduced by molecular oxygen (O_2) absorption. Heating of the atmosphere by sunlight will affect the scale height of N_2 more than of O_2 , possibly causing a relatively higher intensity in the sunlit Northern Hemisphere. This means that the observed differences in the dusk sector could be underestimated, and the difference at dawn may be exaggerated. Another difference is that the observed emissions decrease with increasing mean electron energy, but this effect is stronger for the O I line (VIS). This means that high electron energy could contribute to the asymmetry at dusk. However, the large relative difference observed at 6 MLT, combined with the high absolute intensity in the north, >5.2 kR in Fig. 1a (with a peak intensity of 10 kR), leads us to conclude that the observed asymmetry is not an instrumental effect.

Received 17 March; accepted 21 May 2009.

- Østgaard, N. et al. Observations and model predictions of auroral substorm asymmetries in the conjugate hemispheres. *Geophys. Res. Lett.* **32**, L05111, doi:10.1029/2004GL022166 (2005).
- Østgaard, N., Mende, S. B., Frey, H. U., Sigwarth, J. B., Aasnes, A. & Weygand, J. Auroral conjugacy studies based on global imaging. *J. Atmos. Solar-Terr. Phys.* (special issue) **69**, 249–255 (2007).
- Newell, P. T., Meng, C.-I. & Lyons, K. M. Suppression of discrete aurorae by sunlight. *Nature* **381**, 766–767 (1996).
- Liou, K., Newell, P. T. & Meng, C.-I. Brittnacher, M. & Parks, G. Seasonal effects on auroral particle acceleration and precipitation. *J. Geophys. Res.* **106**, 5531–5542 (2001).
- Richmond, A. D. & Roble, R. G. Electrodynamic effects of thermospheric winds from the NCAR thermospheric general circulation model. *J. Geophys. Res.* **92**, 12365–12376 (1987).
- Benkevich, L., Lyatsky, W. & Cogger, L. L. Field-aligned currents between the conjugate hemispheres. *J. Geophys. Res.* **105**, 27727–27737 (2000).
- Mende, S. B. et al. Far ultraviolet imaging from the IMAGE spacecraft. 2. Wideband FUV imaging. *Space Sci. Rev.* **91**, 271–285 (2000).
- Frank, L. A. et al. The Visible Imaging System (VIS) for the Polar spacecraft. *Space Sci. Rev.* **71**, 297–328 (1995).
- Stenbaek-Nielsen, H. C., Wescott, T. N., Davis, T. N. & Peterson, R. W. Auroral intensity differences at conjugate points. *J. Geophys. Res.* **78**, 659–671 (1973).
- Fillingim, M. O., Parks, G. K., Frey, H. U., Immel, T. J. & Mende, S. B. Hemispheric asymmetry of the afternoon electron aurora. *Geophys. Res. Lett.* **32**, L03113, doi:10.1029/2004GL021635 (2005).
- Stubbs, T. J. & Vondrak, R. R. Østgaard, N., Sigwarth, J. B. & Frank, L. A. Simultaneous observations of the auroral oval in both hemispheres under varying conditions. *Geophys. Res. Lett.* **32**, L03103, doi:10.1029/2004GL021199 (2005).
- Heppner, J. P. & Meynard, N. C. Empirical high-latitude electric field models. *J. Geophys. Res.* **92**, 4467–4489 (1987).
- Cowley, S. W. H. Asymmetry effects associated with the X-component of the IMF in a magnetically open magnetosphere. *Planet. Space Sci.* **29**, 809–818 (1981).
- Sato, N., Nagaoka, T., Hashimoto, K. & Saemundsson, T. Conjugacy of isolated auroral arcs and non-conjugate auroral break-ups. *J. Geophys. Res.* **103**, 11641–11652 (1998).
- Lyatckaya, S., Lyatcky, W. & Khazanov, G. V. Relationship between substorm activity and magnetic disturbances in the two polar caps. *Geophys. Res. Lett.* **35**, L20104, doi:10.1029/2008GL035187 (2008).
- Frank, L. A. & Sigwarth, J. B. Simultaneous images of the northern and southern auroras from the Polar spacecraft: an auroral substorm. *J. Geophys. Res.* **108**, 8015, doi:10.1029/2002JA009356 (2003).
- Frey, H. U. et al. Summary of quantitative interpretation of IMAGE Far Ultraviolet auroral data. *Space Sci. Rev.* **109**, 255–283 (2003).

Acknowledgements We are indebted to the IMAGE and Polar teams for the design and successful operations of the two missions. We especially thank S. B. Mende for the use of IMAGE FUV WIC data and J. B. Sigwarth for the use of Polar VIS Earth Camera data. This study was supported by the Norwegian Research Council, through the IPY-ICESTAR project 176045/S30

Author Contributions N.Ø is responsible for the project planning, K.M.L. is responsible for the data processing. The data were identified and analysed by K.M.L. and N.Ø. The manuscript was written by K.M.L. and N.Ø.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to N.Ø. (nikolai.ostgaard@uib.no).

LETTERS

State-independent experimental test of quantum contextuality

G. Kirchmair^{1,2}, F. Zähringer^{1,2}, R. Gerritsma^{1,2}, M. Kleinmann¹, O. Gühne^{1,3}, A. Cabello⁴, R. Blatt^{1,2} & C. F. Roos^{1,2}

The question of whether quantum phenomena can be explained by classical models with hidden variables is the subject of a long-lasting debate^{1,2}. In 1964, Bell showed that certain types of classical models cannot explain the quantum mechanical predictions for specific states of distant particles, and some types of hidden variable models^{3–9} have been experimentally ruled out. An intuitive feature of classical models is non-contextuality: the property that any measurement has a value independent of other compatible measurements being carried out at the same time. However, a theorem derived by Kochen, Specker and Bell^{10–12} shows that non-contextuality is in conflict with quantum mechanics. The conflict resides in the structure of the theory and is independent of the properties of special states. It has been debated whether the Kochen–Specker theorem could be experimentally tested at all^{13,14}. First tests of quantum contextuality have been proposed only recently, and undertaken with photons^{15,16} and neutrons^{17,18}. But these tests required the generation of special quantum states and left various loopholes open. Here we perform an experiment with trapped ions that demonstrates a state-independent conflict with non-contextuality. The experiment is not subject to the detection loophole and we show that, despite imperfections and possible measurement disturbances, our results cannot be explained in non-contextual terms.

Hidden variable models assert that the result $v(A)$ of measuring the observable A on an individual quantum system is predetermined by a hidden variable λ . Two observables A and B are mutually compatible if the result of A does not depend on whether B is measured before, after or simultaneously with A and vice versa. Non-contextuality is the property of a hidden variable model that the value $v(A)$ is predetermined regardless of which other compatible observable is measured jointly with A . Kochen and Specker showed that the assumption of non-contextuality cannot be reconciled with quantum mechanics. A considerable simplification of the original Kochen–Specker argument by Mermin and Peres^{19,20} uses a 3×3 square of observables A_{ij} with possible outcomes $v(A_{ij}) = \pm 1$, where the observables in each row or column are mutually compatible. Considering the products of rows $R_k = v(A_{k1})v(A_{k2})v(A_{k3})$ and columns $C_k = v(A_{1k})v(A_{2k})v(A_{3k})$, the total product would be $\prod_{k=1,2,3} R_k C_k = 1$, as any $v(A_{ij})$ appears twice in the total product.

In quantum mechanics, however, one can take a four-level quantum system, for instance two spin- $\frac{1}{2}$ -particles, and the following array of observables:

$$\begin{array}{lll} A_{11} = \sigma_z^{(1)} & A_{12} = \sigma_x^{(2)} & A_{13} = \sigma_z^{(1)} \otimes \sigma_z^{(2)} \\ A_{21} = \sigma_x^{(2)} & A_{22} = \sigma_x^{(1)} & A_{23} = \sigma_x^{(1)} \otimes \sigma_x^{(2)} \\ A_{31} = \sigma_z^{(1)} \otimes \sigma_x^{(2)} & A_{32} = \sigma_x^{(1)} \otimes \sigma_z^{(2)} & A_{33} = \sigma_y^{(1)} \otimes \sigma_y^{(2)} \end{array} \quad (1)$$

Here $\sigma_i^{(k)}$ denotes the Pauli matrix acting on the k th particle, and all the observables have the outcomes ± 1 . Moreover, in each of the rows

or columns of the Mermin–Peres square (1), the observables are mutually commuting and can be measured simultaneously or in any order, hence they are compatible. In any row or column, their measurement product R_k or C_k equals 1, except for the third column where it equals -1 . Therefore, quantum mechanics yields for the product $\prod_{k=1,2,3} R_k C_k$ a value of -1 , in contrast to non-contextual models.

To test this property, it has to be expressed as an inequality because no experiment yields ideal quantum measurements. Recently, it has been shown that the inequality

$$\langle \mathcal{X}_{KS} \rangle = \langle R_1 \rangle + \langle R_2 \rangle + \langle R_3 \rangle + \langle C_1 \rangle + \langle C_2 \rangle - \langle C_3 \rangle \leq 4 \quad (2)$$

holds for all non-contextual theories²¹, where $\langle \dots \rangle$ denotes the ensemble average. Quantum mechanics predicts for any state that the mean value in the Kochen–Specker inequality is $\langle \mathcal{X}_{KS} \rangle = 6$, thereby violating inequality (2). For an experimental test, an ensemble of quantum states Ψ needs to be prepared and each realization subjected to the measurement of one of the possible sets of compatible observables. Here, it is of utmost importance that all measurements of A_{ij} are context-independent²¹, that is, A_{ij} must be detected with a quantum non-demolition (QND) measurement that provides no information whatsoever about any other compatible observable.

Experiments processing quantum information with trapped ions²² are particularly well-suited for this purpose, as arbitrary two-qubit quantum states Ψ can be deterministically generated by laser–ion interactions and measured with near-unit efficiency⁶. Two ionic energy levels are defined to represent the qubit basis states $|\uparrow\rangle$ and $|\downarrow\rangle$, which are eigenstates of the observable σ_z . The qubit is measured by electron shelving²² projecting onto $|\uparrow\rangle$ or $|\downarrow\rangle$. Measurement of any other observable A_{ij} is reduced to detecting $\sigma_z^{(k)}$, $k = 1$ or 2 , by applying a suitable unitary transformation U to the state Ψ before measuring $\sigma_z^{(k)}$, and its inverse U^\dagger after the measurement (see Fig. 1 and Methods). With these basic tools, any set of observables can be sequentially measured in an arbitrary temporal order.

For the experiment, a pair of ${}^{40}\text{Ca}^+$ ions is trapped in a linear Paul trap with axial and radial vibrational frequencies of $w_{\text{ax}} = (2\pi) 1.465$ MHz and $w_r \approx (2\pi) 3.4$ MHz, respectively, and Doppler-cooled by exciting the $S_{1/2} \leftrightarrow P_{1/2}$ and $P_{1/2} \leftrightarrow D_{3/2}$ dipole transitions. Optical pumping initializes an ion with a fidelity of 99.5% to the qubit state $|\downarrow\rangle \equiv |S_{1/2}, m = 1/2\rangle$, the second qubit state being $|\uparrow\rangle \equiv |D_{5/2}, m = 3/2\rangle$. The qubit is coherently manipulated²³ by an ultrastable, narrow-band laser coherently exciting the $S_{1/2} \leftrightarrow D_{5/2}$ quadrupole transition in a magnetic field of $B = 4$ G. Single-qubit light-shift gates $U_z^{(1)}(\theta) = \exp(-i\frac{\theta}{2}\sigma_z^{(1)})$ are realized by an off-resonant beam impinging on ion 1 with a beam waist of $3 \mu\text{m}$ and a k -vector perpendicular to the ion string. A second beam, illuminating both ions

¹Institut für Quantenoptik und Quanteninformation, Österreichische Akademie der Wissenschaften, Otto-Hittmair-Platz 1, A-6020 Innsbruck, Austria. ²Institut für Experimentalphysik, ³Institut für theoretische Physik, Universität Innsbruck, Technikerstr. 25, A-6020 Innsbruck, Austria. ⁴Departamento de Física Aplicada II, Universidad de Sevilla, E-41012 Sevilla, Spain.

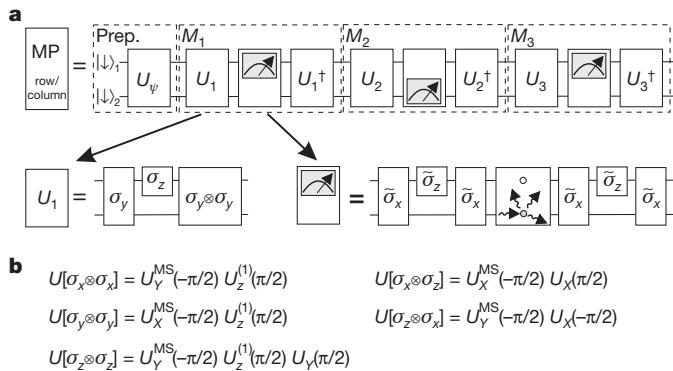


Figure 1 | Experimental measurement scheme. **a**, For the measurement of the j th row (column) of the Mermin–Peres (MP) square (1), a quantum state is prepared (Prep.) on which three consecutive QND measurements M_k , $k = 1, 2, 3$, are performed measuring the observables $A_{jk}(A_{kj})$. Each measurement consists of a composite unitary operation U_k that maps the observable of interest onto one of the single-qubit observables $\sigma_z^{(1)}$ or $\sigma_z^{(2)}$, which are measured by fluorescence detection. The unitary operations U are synthesized from single-qubit and maximally entangling gates. In the lower row, σ_i , $\tilde{\sigma}_i$ symbolize, respectively, the Hamiltonian acting on the qubit and on the $D_{5/2}$ Zeeman subspace spanned by $\{| \downarrow \rangle, | a \rangle\}$ which is used for hiding one ion’s quantum state from fluorescence light during detection³⁰. **b**, All mapping operations U_k used for measuring the five two-qubit spin correlations $\sigma_i^{(1)} \otimes \sigma_j^{(2)}$ require an entangling gate. Here, we list the gate decompositions of $U[\sigma_i^{(1)} \otimes \sigma_j^{(2)}]$ used in the experiments, where $U_X(\theta) \equiv U(\theta, \phi = 0)$ and $U_Y(\theta) \equiv U(\theta, \phi = \pi/2)$.

with equal strength at an angle of 45° with respect to the ion crystal, serves to carry out gate operations that are symmetric under qubit exchange. Collective single-qubit gates $U(\theta, \phi) = \exp(-i\frac{\theta}{2}(\sigma_\phi^{(1)} + \sigma_\phi^{(2)}))$, where $\sigma_\phi^{(k)} = \cos(\phi)\sigma_x^{(k)} + \sin(\phi)\sigma_y^{(k)}$, are realized by resonantly exciting the qubit transition and controlling the phase ϕ of the laser light. If instead a bichromatic light field near-resonant with the upper and lower sideband transitions of the axial centre-of-mass (COM) mode is used, a Mølmer–Sørensen gate²⁴ $U^{\text{MS}}(\theta, \phi) = \exp(-i\frac{\theta}{2}\sigma_\phi^{(1)} \otimes \sigma_\phi^{(2)})$ is implemented^{23,25}. We achieve a maximally entangling gate ($\theta = \pi/2$) capable of mapping $|\downarrow\downarrow\rangle \rightarrow |\downarrow\downarrow\rangle + ie^{-i2\phi}|\uparrow\uparrow\rangle$ with a fidelity of about 98%, even for Doppler-cooled ions in a thermal state with an average of $\bar{n}_{\text{ax, COM}} \approx 18$ vibrational quanta. This property is of crucial importance, as the experiment demands gate operations subsequent to quantum state detection by fluorescence measurements which do not preserve the motional quantum number. The set of elementary gates²⁶ $\{U_z^{(1)}(\theta), U(\theta, \phi), U^{\text{MS}}(\theta, \phi)\}$ is sufficient to construct the two-qubit unitary operations needed for creating various input states Ψ and mapping the observables A_{ij} to $\sigma_z^{(k)}$ for read-out (see Methods).

Equipped with these tools, we create the singlet state $\Psi = (|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle)/\sqrt{2}$ by applying the gates $U_z^{(1)}(\pi) U_z^{(\frac{\pi}{2}, \frac{3\pi}{4})} U^{\text{MS}}(\frac{\pi}{2}, 0)$ to the initial state $|\downarrow\downarrow\rangle$ and measure consecutively the three observables of a row or column of the Mermin–Peres square. The results obtained for a total of 6,600 copies of Ψ are visualized in Fig. 2. The three upper panels (Fig. 2a–c) show the distribution of measurement results $\{v(A_{11}), v(A_{12}), v(A_{13})\}$ and their products for the observables appearing in the rows of the Mermin–Peres square (1), and the three lower panels (Fig. 2d–f) show the corresponding results for the columns of the square. Figure 2f demonstrates that Ψ is a common eigenstate of the observables $\sigma_x^{(1)} \otimes \sigma_x^{(2)}$, $\sigma_y^{(1)} \otimes \sigma_y^{(2)}$, $\sigma_z^{(1)} \otimes \sigma_z^{(2)}$, as only one of the spheres has a considerable size. Five of the correlations have a value close to +1 whereas $\langle C_3 \rangle = -0.91(1)$. By adding them up and subtracting $\langle C_3 \rangle$, we find a value of $\langle \mathcal{X}_{\text{KS}} \rangle = 5.46(4) > 4$, thus violating equation (2). Here and elsewhere in the paper, errors are due to slowly fluctuating experimental control parameters.

To test the prediction of a state-independent violation, we repeated the experiment for nine other quantum states of different

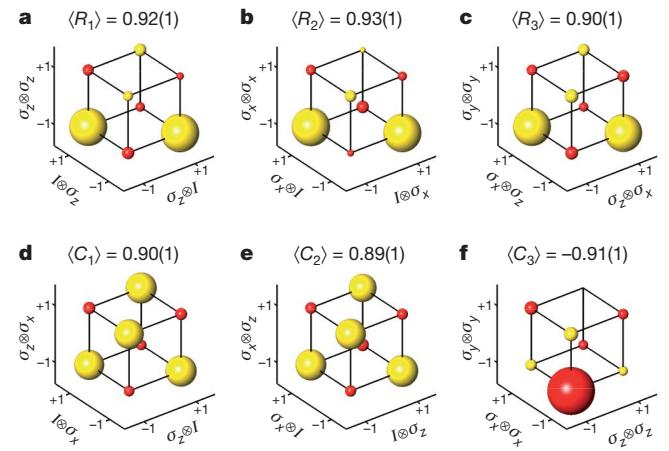


Figure 2 | Measurement correlations for the singlet state. **a**, Visualization of the consecutive measurement of the three observables $A_{11} = \sigma_z^{(1)}$, $A_{12} = \sigma_z^{(2)}$, $A_{13} = \sigma_z^{(1)} \otimes \sigma_z^{(2)}$ corresponding to row 1 of the Mermin–Peres square. The measurement is carried out on 1,100 preparations of the singlet state. The volume of the spheres on each corner of the cube represents the relative frequency of finding the measurement outcome $\{v_1, v_2, v_3\}$, $v_i \in \{\pm 1\}$. The exact probabilities for finding $\{v_1, v_2, v_3\}$ are given in Supplementary Information. The colour of the sphere indicates whether $v_1 v_2 v_3 = +1$ (yellow) or -1 (red). The average of the measurement product $\langle R_1 \rangle$ is given at the top. **b–f**, Measurements of the remaining rows (**b**, **c**) or columns (**d–f**) of the Mermin–Peres square. Errors given are 1σ .

purity and entanglement. Figure 3 shows that a state-independent violation of the Kochen–Specker inequality indeed occurs, with $\langle \mathcal{X}_{\text{KS}} \rangle$ ranging from 5.23(5) to 5.46(4). We also checked that a violation of inequality (2) occurs irrespective of the temporal order of the measurement triples. Figure 4 shows the results for all possible permutations of the rows and columns of (1), based on 39,600 realizations of the singlet state. When combining the correlation results for the 36 possible permutations of operator orderings in (1), we find that all values for the Kochen–Specker inequality $\langle \mathcal{X}_{\text{KS}} \rangle$ fall within the range from 5.22 to 5.49. Because of experimental imperfections, the experimental violation of the Kochen–Specker inequality falls short of the quantum-mechanical prediction. The dominating error

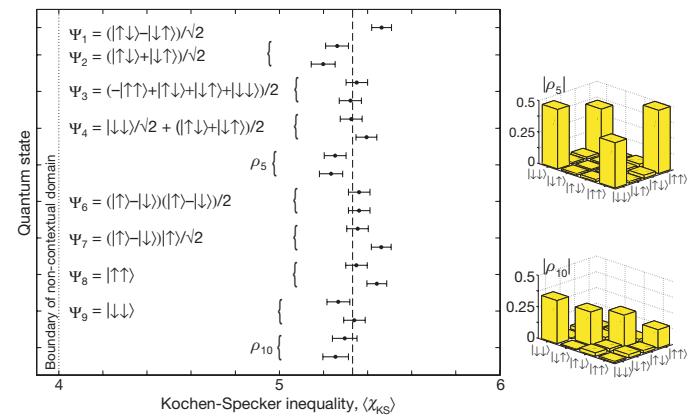


Figure 3 | State-independence of the Kochen–Specker inequality. The Kochen–Specker inequality was tested for ten different quantum states, including maximally entangled (Ψ_1 – Ψ_3), partially entangled (Ψ_4) and separable (Ψ_6 – Ψ_9) almost pure states, as well as an entangled mixed state (ρ_5) and an almost completely mixed state (ρ_{10}). All states are analysed by quantum state tomography, which yields for the experimentally produced states Ψ_1 – Ψ_4 , Ψ_6 – Ψ_9 , an average fidelity of 97(2)%. For ρ_5 and ρ_{10} , the tomographic reconstruction is shown at right. For all states, we obtain a violation of inequality (2) which demonstrates its state-independent character, the vertical dashed line indicating the average value of $\langle \mathcal{X}_{\text{KS}} \rangle$. Error bars, 1σ (6,600 state realizations per data point).

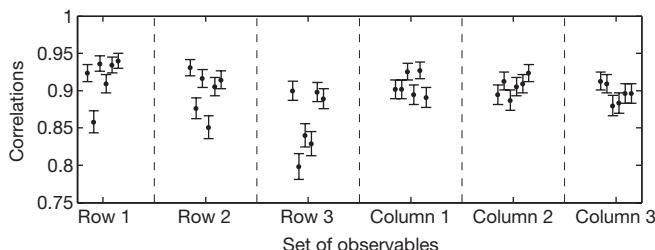


Figure 4 | Permutation within rows and columns of the Mermin–Peres square. As the three observables of a set are commuting, the temporal order of their measurements should have no influence on the measurement results. The figure shows the measured absolute values of the products of observables for any of the six possible permutations. The scatter in the experimental data is caused by experimental imperfections that affect different permutations differently. For the measurements shown here, 39,600 copies of the singlet state were used in total. Error bars, 1σ.

sources are imperfect unitary operations, in particular the entangling gates applied up to six times in a single experimental run.

All experimental tests of hidden-variable theories are subject to various possible loopholes. In our experiment, the detection loophole does not have a role, as the states of the ions are detected with near-perfect efficiency. From the point of view of a hidden variable theory, objections can still be made: in the experiment, owing to imperfect mapping operations, the observables are not perfectly compatible and as the observables are measured sequentially, it may be that the hidden variables are disturbed during the sequence of measurements, weakening the necessity to assign to any observable a fixed value independently of the context.

Nevertheless, it is possible to derive inequalities for classical non-contextual models, wherein the hidden variables are disturbed during the measurement process (see Methods). More specifically, it can be proved that the probabilities of measurement outcomes for the disturbed hidden variables then obey the inequality:

$$\begin{aligned} \langle \mathcal{X}_{\text{DHV}} \rangle &= \langle A_{12}A_{13} \rangle + \langle A_{22}A_{23} \rangle + \langle A_{12}A_{22} \rangle - \langle A_{13}A_{23} \rangle \\ &\quad - 2p^{\text{err}}[A_{13}A_{12}A_{13}] - 2p^{\text{err}}[A_{23}A_{22}A_{23}] \\ &\quad - 2p^{\text{err}}[A_{22}A_{12}A_{22}] - 2p^{\text{err}}[A_{23}A_{13}A_{23}] \leq 2 \end{aligned} \quad (3)$$

Here, $\langle A_i A_j \rangle = \langle \nu(A_i) \nu(A_j) \rangle$ denotes the ensemble average if A_i is measured before A_j , and $p^{\text{err}}[A_i A_j A_i]$ denotes the probability that measuring A_j introduces a change in the value of A_i if the sequence $A_i A_j A_i$ is measured. To test this inequality, we prepared the state $\Psi \propto |\uparrow\uparrow\rangle + i|\downarrow\uparrow\rangle + \gamma|\uparrow\downarrow\rangle + i|\downarrow\downarrow\rangle$ where $\gamma = \sqrt{2} - 1$, and measured the Mermin–Peres square with σ_y and σ_z exchanged. We find for the whole square the value $\langle \mathcal{X}_{\text{KS}} \rangle = 5.22(10) > 4$, and for equation (3) the value $\langle \mathcal{X}_{\text{DHV}} \rangle = 2.23(5) > 2$. This proves that even disturbances of the hidden variables during not perfectly compatible measurements cannot explain the given experimental data.

In principle, our analysis of measurement disturbances and dynamical hidden variable models can be extended to the full Mermin–Peres square, but the experimental techniques have to be improved to find a violation there. Our findings already show unambiguously that the experimentally observed phenomena cannot be described by non-contextual models. Remarkably, the experimental observation here of counter-intuitive quantum predictions did not require the preparation of specific entangled states with non-local correlations. We expect this result to stimulate new applications of quantum contextuality for quantum information processing^{27–29}.

METHODS SUMMARY

In quantum physics experiments, there is often a preferred basis for carrying out measurements. In the case of trapped ion qubits, measurements projecting onto eigenstates of σ_z are easily carried out by a fluorescence detection that couples only one of the qubit states $|\uparrow\rangle, |\downarrow\rangle$ to an excited state that rapidly decays back to the same qubit state. In the two-ion experiment described in this paper, the

observables of interest, $\sigma_i^{(1)} \equiv \sigma_i \otimes I$, $\sigma_j^{(2)} \equiv I \otimes \sigma_j$ and $\sigma_i^{(1)} \otimes \sigma_j^{(2)}$, all have two doubly-degenerate eigenvalues. Therefore, these observables can be measured by unitarily mapping them to $\sigma_i^{(k)}$, $k = 1$ or 2, measuring $\sigma_i^{(k)}$ and applying the inverse mapping operation. For spin correlations $\sigma_i \otimes \sigma_j$, the mapping requires the application of entangling two-qubit gates.

The inequality (3) holds for hidden variable models, which assign definite values to any observable and allow the probability distribution of the hidden variables to change during sequential measurements. It is assumed that (on average) a measurement of an observable A followed by a measurement of an observable B disturbs the predetermined value of A more strongly than a measurement of B alone. Then, from a Clauser–Horne–Shimony–Holt-type inequality²¹, the inequality (3) follows.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 8 April; accepted 26 May 2009.

- Einstein, A., Podolsky, B. & Rosen, N. Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.* **47**, 777–780 (1935).
- Bell, J. S. On the Einstein–Podolsky–Rosen paradox. *Physics* **1**, 195–200 (1964).
- Aspect, A., Dalibard, J. & Roger, G. Experimental test of Bell's inequalities using time-varying analyzers. *Phys. Rev. Lett.* **49**, 1804–1807 (1982).
- Tittel, W., Brendel, J., Zbinden, H. & Gisin, N. Violation of Bell inequalities by photons more than 10 km apart. *Phys. Rev. Lett.* **81**, 3563–3566 (1998).
- Weis, G., Jennewein, T., Simon, C., Weinfurter, H. & Zeilinger, A. Violation of Bell's inequality under strict Einstein locality conditions. *Phys. Rev. Lett.* **81**, 5039–5043 (1998).
- Rowe, M. A. et al. Experimental violation of a Bell's inequality with efficient detection. *Nature* **409**, 791–794 (2001).
- Gröblacher, S. et al. An experimental test of non-local realism. *Nature* **446**, 871–875 (2007).
- Branciard, C. et al. Testing quantum correlations versus single-particle properties within Leggett's model and beyond. *Nature Phys.* **4**, 681–685 (2008).
- Matsukevich, D. N., Maunz, P., Moehrung, D. L., Olmschenk, S. & Monroe, C. Bell inequality violation with two remote atomic qubits. *Phys. Rev. Lett.* **100**, 150404 (2008).
- Specker, E. Die Logik nicht gleichzeitig entscheidbarer Aussagen. *Dialectica* **14**, 239–246 (1960).
- Bell, J. S. On the problem of hidden variables in quantum mechanics. *Rev. Mod. Phys.* **38**, 447–452 (1966).
- Kochen, S. & Specker, E. P. The problem of hidden variables in quantum mechanics. *J. Math. Mech.* **17**, 59–87 (1967).
- Cabello, A. & García-Alcaine, G. Proposed experimental tests of the Bell–Kochen–Specker theorem. *Phys. Rev. Lett.* **80**, 1797–1799 (1998).
- Meyer, D. A. Finite precision measurement nullifies the Kochen–Specker theorem. *Phys. Rev. Lett.* **83**, 3751–3754 (1999).
- Michler, M., Weinfurter, H. & Żukowski, M. Experiments towards falsification of noncontextual hidden variable theories. *Phys. Rev. Lett.* **84**, 5457–5461 (2000).
- Huang, Y.-F., Li, C.-F., Zhang, Y.-S., Pan, J.-W. & Guo, G.-C. Experimental test of the Kochen–Specker theorem with single photons. *Phys. Rev. Lett.* **90**, 250401 (2003).
- Hasegawa, Y., Loidl, R., Badurek, G., Baron, M. & Rauch, H. Quantum contextuality in a single-neutron optical experiment. *Phys. Rev. Lett.* **97**, 230401 (2006).
- Bartosik, H. et al. Experimental test of quantum contextuality in neutron interferometry. Preprint at <http://arXiv.org/abs/0904.4576> (2009).
- Peres, A. Incompatible results of quantum measurements. *Phys. Lett. A* **151**, 107–108 (1990).
- Mermin, N. D. Simple unified form for the major no-hidden-variables theorems. *Phys. Rev. Lett.* **65**, 3373–3376 (1990).
- Cabello, A. Experimentally testable state-independent quantum contextuality. *Phys. Rev. Lett.* **101**, 210401 (2008).
- Häffner, H., Roos, C. F. & Blatt, R. Quantum computing with trapped ions. *Phys. Rep.* **469**, 155–203 (2008).
- Kirchmair, G. et al. Deterministic entanglement of ions in thermal states of motion. *N. J. Phys.* **11**, 023002 (2009).
- Mølmer, K. & Sørensen, A. Multiparticle entanglement of hot trapped ions. *Phys. Rev. Lett.* **82**, 1835–1838 (1999).
- Benhelm, J., Kirchmair, G., Roos, C. F. & Blatt, R. Towards fault-tolerant quantum computing with trapped ions. *Nature Phys.* **4**, 463–466 (2008).
- Nebendahl, V., Häffner, H. & Roos, C. F. Optimal control of entangling operations for trapped-ion quantum computing. *Phys. Rev. A* **79**, 012312 (2009).
- Bechmann-Pasquinucci, H. & Peres, A. Quantum cryptography with 3-state systems. *Phys. Rev. Lett.* **85**, 3313–3316 (2000).
- Galvão, E. F. *Foundations of Quantum Theory and Quantum Information Applications*. Ph.D. thesis, Oxford Univ. (2002).
- Spekkens, R. W., Bužek, D. H., Keehn, A. J., Toner, B. & Pryde, G. J. Preparation contextuality powers parity-oblivious multiplexing. *Phys. Rev. Lett.* **102**, 010401 (2009).
- Roos, C. F. et al. Control and measurement of three-qubit entangled states. *Science* **304**, 1478–1480 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge support by the Austrian Science Fund (FWF), by the European Commission (SCALA, OLAQUI and QICS networks, and the Marie-Curie programme), by the Institut für Quanteninformation GmbH, by the Spanish MCI Project FIS2008-05596 and the Junta de Andalucía Excellence Project P06-FQM-02243. This material is based on work supported in part by IARPA.

Author Contributions G.K., F.Z. and R.G. performed the experiment and partially analysed the data; M.K., O.G. and A.C. provided the theoretical part and the modelling of imperfect measurements; C.F.R. conceived the experiment and analysed the data; R.B., G.K., F.Z., R.G. and C.F.R. contributed to the experimental set-up; and all authors co-wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to C.F.R. (Christian.Roos@uibk.ac.at).

METHODS

Quantum state detection. The quantum state of a single ion is detected by illuminating both ions with light near-resonant with the $S_{1/2} \leftrightarrow P_{1/2}$ and $P_{1/2} \leftrightarrow D_{3/2}$ transitions for 250 μs . To prevent the quantum information of the other ion from being read out, the quantum state of the qubit that is not to be detected is hidden in the $D_{5/2}$ -Zeeman state manifold by a composite π -pulse transferring the $|\downarrow\downarrow\rangle$ state's population to the auxiliary state $|a\rangle \equiv |D_{5/2}, m=5/2\rangle$ before fluorescence detection. After the detection, the qubit is restored in the state space $\{|\uparrow\uparrow\rangle, |\downarrow\downarrow\rangle\}$ by another composite π -pulse. The parameters of the read-out lasers are set such that it keeps the axial COM-mode in the Lamb-Dicke regime²² with $\bar{n}_{\text{COM}} \approx 18$. By combining the counts of two photomultipliers, we observe a Poissonian distribution of photon counts in the detection window with average count numbers $n_{\uparrow} \approx 7.8$ and $n_{\downarrow} \approx 0.07$ in the bright and dark state, respectively. Setting the discrimination threshold to 1.5 counts, the conditional probabilities for wrong quantum state assignments amount to $p(\uparrow|\downarrow) \approx 0.24\%$ and $p(\downarrow|\uparrow) \approx 0.39\%$. At the end of the detection interval, the ion is optically pumped on the $S_{1/2} \leftrightarrow P_{1/2}$ to prevent leakage of population from the qubit level $|\downarrow\downarrow\rangle$ to the state $|S_{1/2}, m=-1/2\rangle$.

QND measurements of spin correlations. Quantum non-demolition measurements of observables A_{ij} that measure spin correlations are carried out by mapping the subspace $\mathcal{H}_{A_{ij}}^+ = \{\psi | A_{ij}\psi = \psi\}$ onto the subspace $\mathcal{H}_{\sigma_z^{(2)}}^+ = \{\psi | \sigma_z^{(2)}\psi = \psi\}$ before the fluorescence measurement of $\sigma_z^{(2)}$. This is achieved by applying a unitary state transformation U_{ij} satisfying $A_{ij} = U_{ij}^\dagger \sigma_z^{(2)} U_{ij}$ to the two-qubit state of interest. To decompose U_{ij} into the elementary gate operations available in our set-up, we use a gradient-ascent based numerical search routine²⁶. After measurement of $\sigma_z^{(2)}$, the inverse operation U_{ij}^\dagger completes the QND measurement.

Modelling imperfect measurements. To deal with the case of imperfect measurements from a hidden variable viewpoint, let us assume that there is a hidden variable model that determines jointly the probabilities of the results of all sequences of measurements. Such probabilities are written as $p[A^{(1)+}, B^{(2)-}; AB]$ and so on, denoting the probability for the result $A^{(1)} = +1$ and $B^{(2)} = -1$ when measuring first A and then B . Then the probabilities fulfil:

$$p\left[\left(A^{(1)+}; A\right) \text{and} \left(B^{(1)+}; B\right)\right] \leq p\left[A^{(1)+}, B^{(2)+}; AB\right] + \\ p\left[\left(B^{(1)+}; B\right) \text{and} \left(B^{(2)-}; AB\right)\right] \quad (4)$$

This inequality holds because if the hidden variable λ is such that it contributes to $p[A^{(1)+}; A]$ and $(B^{(1)+}; B)$, then either the value of B stays the same when measuring A and λ contributes to $p[A^{(1)+}, B^{(2)+}; AB]$, or it is flipped and it contributes to $p[(B^{(1)+}; B) \text{ and } (B^{(2)-}; AB)]$. The term $p[A^{(1)+}, B^{(2)+}; AB]$ is directly measurable, and we estimate the other term by assuming:

$$p\left[\left(B^{(1)+}; B\right) \text{ and} \left(B^{(2)-}; AB\right)\right] \leq p\left[B^{(1)+}, B^{(3)-}; BAB\right] \quad (5)$$

This inequality means that the disturbance of a predetermined value of B caused by the measurement of B and A should be not smaller than the disturbance due to measurement of A alone, as the former includes additional experimental procedures compared with the latter. The probability $p[B^{(1)+}, B^{(3)-}; BAB]$ is experimentally accessible, and combining equation (4) and (5) we obtain a measurable upper bound on $p[A^{(1)+}; A]$ and $(B^{(1)+}; B)$. Then, the Clauser-Horne-Shimony-Holt-type inequality²¹ $\langle AB \rangle + \langle CD \rangle + \langle AC \rangle - \langle BD \rangle \leq 2$ holds, where the correlations like $\langle AB \rangle$ are obtained from the probabilities $p[(A^{(1)+}; A) \text{ and } (B^{(1)+}; B)]$ and so on. Using the above bounds and the notation $p^{\text{err}}[BAB] = p[B^{(1)+}, B^{(3)-}; BAB] + p[B^{(1)-}, B^{(3)+}; BAB]$, one arrives at the inequality (3) for sequences of measurements.

The probabilities $p^{\text{err}}[BAB]$ quantify the compatibility of the implemented observables. For the measurements in inequality (3) they are found to be 0.047(17) on average. For the full Mermin–Peres square, a similar analysis would need error probabilities of the kind $p^{\text{err}}[BAB]$ and $p^{\text{err}}[CABC]$ to deal, for example, with the term $\langle ABC \rangle$. For a measured value of $\langle \mathcal{X}_{\text{KS}} \rangle = 5.46$, they would have to be on average smaller than 0.03 in order to find a violation of an inequality similar to (3). This error level, however, is more difficult to achieve as measurement results have to be correlated that involve subsequent measurement of up to four observables.

LETTERS

Near-field focusing and magnification through self-assembled nanoscale spherical lenses

Ju Young Lee^{1*}, Byung Hee Hong^{1,2*}, Woo Youn Kim¹, Seung Kyu Min¹, Yukyung Kim¹, Mikhail V. Jouravlev¹, Ranojoy Bose³, Keun Soo Kim², In-Chul Hwang¹, Laura J. Kaufman⁴, Chee Wei Wong³, Philip Kim⁵ & Kwang S. Kim¹

It is well known that a lens-based far-field optical microscope cannot resolve two objects beyond Abbe's diffraction limit. Recently, it has been demonstrated that this limit can be overcome by lensing effects driven by surface-plasmon excitation^{1–3}, and by fluorescence microscopy driven by molecular excitation⁴. However, the resolution obtained using geometrical lens-based optics without such excitation schemes remains limited by Abbe's law even when using the immersion technique⁵, which enhances the resolution by increasing the refractive indices of immersion liquids. As for submicrometre-scale or nanoscale objects, standard geometrical optics fails for visible light because the interactions of such objects with light waves are described inevitably by near-field optics⁶. Here we report near-field high resolution by nanoscale spherical lenses that are self-assembled by bottom-up integration⁷ of organic molecules. These nanolenses, in contrast to geometrical optics lenses, exhibit curvilinear trajectories of light, resulting in remarkably short near-field focal lengths. This in turn results in near-field magnification that is able to resolve features beyond the diffraction limit. Such spherical nanolenses provide new pathways for lens-based near-field focusing and high-resolution optical imaging at very low intensities, which are useful for bio-imaging, near-field lithography, optical memory storage, light harvesting, spectral signal enhancing, and optical nano-sensing.

Miniaturized lenses are often found in biological systems^{8,9}, and have been widely used for optical microelectromechanical systems¹⁰. Despite numerous studies of miniaturized lenses^{11–14}, no serious studies have been undertaken of lenses for subwavelength nano-optics. To this end, we fabricate well-defined nanoscale lenses of calix[4]hydroquinone (CHQ), which is composed of four *p*-hydroquinone subunits with eight hydroxyl groups¹⁵. Intermolecular short hydrogen-bonding¹⁶ and π - π stacking interactions¹⁷ are very useful forces for self-assembling supramolecular CHQ nanostructures¹⁸. One class of such self-assembled nanostructures is represented by sphere-derived shapes of diameter 50 nm to 3 μ m, as seen in scanning electron microscopy (SEM) images (Fig. 1a–c). Dissolving the CHQ monomers in 1:1 water–acetone solution leads to the formation of needle-like CHQ nanotube crystals with infinitely long hydrogen-bonded arrays. As the crystals grown at –14 °C are heated at 40 °C in aqueous environments for a day, CHQ molecules released from the crystals re-assemble into nanospheres.

At the beginning of this process, film-like structures of CHQ cover the surface of the crystals. CHQ molecules released from the surface accumulate in a small volume under the film, leading to the nucleation and growth of two-dimensional disk-shaped structures. Spherical curvatures are then gradually formed, as more of the

released CHQ molecules re-assemble in three dimensions. These are intermediate structures of anisotropically growing spheres (see Supplementary Information A for the mechanism). These 'plano-spherical convex' (PSC) structures—with a spherical face on one side and a flat face on the other side—can be isolated, and these CHQ lenses are stable in air. The negative electron beam resist properties of calixarene-based structures¹⁹ are used for fabrication and positioning of optical devices using these CHQ lenses (Supplementary Information D). Although these experiments are technically sophisticated, the intricately self-assembled nanolenses enable us to study the physics of nanolens optics.

The size distribution of CHQ lenses can be controlled by the time and temperature of the self-assembly process. Typically, PSC lenses

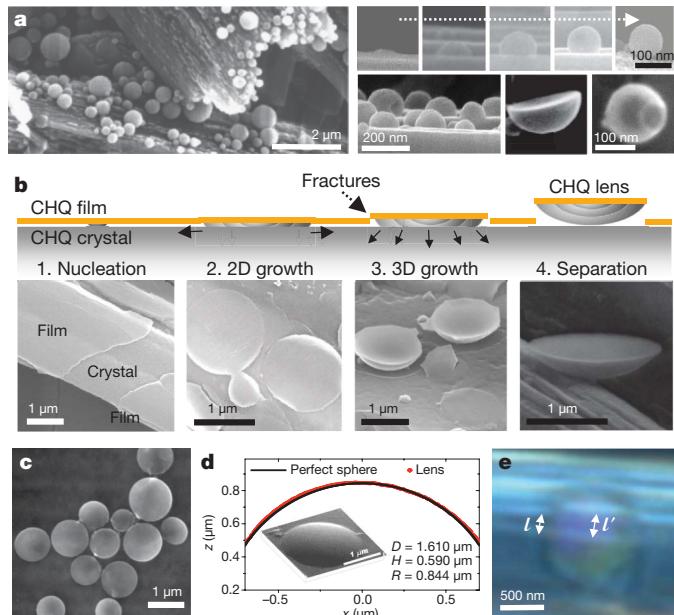


Figure 1 | CHQ plano-spherical convex lenses. **a**, SEM images of growing CHQ nanospheres and their intermediate structures. **b**, Schematic diagrams and SEM images showing the self-assembly of CHQ lenses (see text for details). **c**, SEM image showing various sizes of CHQ lenses separated as an aqueous suspension and drop-dried on a substrate. **d**, AFM profile showing the near perfect spherical face of the lens. Inset, corresponding SEM image. **e**, Optical microscope image of CHQ lenses on a CHQ nanotube crystal, showing the magnification by the lens. The line spacing (l) behind the lens is considerably increased (l'). 500 nm scale bar.

¹Center for Superfunctional Materials, Department of Chemistry, Pohang University of Science and Technology, Hyojadong, Namgu, Pohang 790-784, Korea. ²Department of Chemistry and SKKU Advanced Institute of Nanotechnology, Sungkyunkwan University, Suwon 440-746, Korea. ³Department of Mechanical Engineering, ⁴Department of Chemistry, ⁵Department of Physics, Columbia University, New York, New York 10027, USA.

*These authors contributed equally to this work.

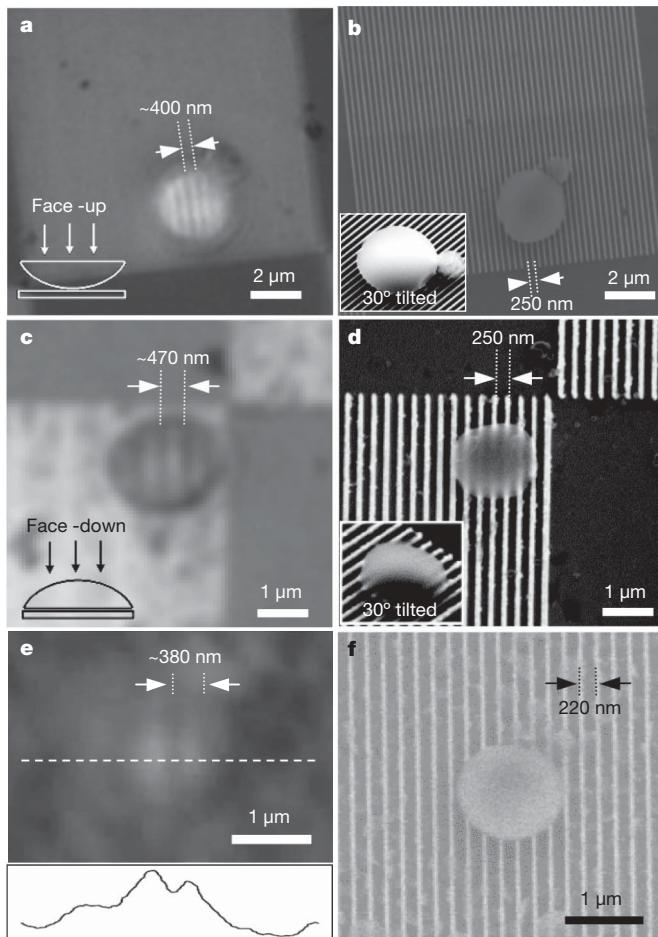


Figure 2 | Optical microscope/SEM images of CHQ lenses on patterned substrates. **a**, Optical microscope image of a face-up lens placed on a glass substrate with Pd stripe patterns (see Supplementary Information C for more images taken with different focus and magnification). **b**, SEM image corresponding to **a**. Inset, image of tilted lens. **c**, Optical microscope image of a face-down lens. **d**, SEM image corresponding to **c**. **e**, Optical microscope image of a face-up lens. Inset, light intensity profile taken from the blue dotted line. **f**, SEM image corresponding to **e**. The sub-diffraction-limit patterns cannot be resolved in conventional optical microscopy, but the magnifying effect through the lens allows the stripe patterns of 250/220 nm spacing to be resolved.

with nanoscale thickness $H < 800$ nm and diameter $D = 0.05\text{--}3 \mu\text{m}$ can be synthesized and separated from the aqueous suspension for further experiments. Surface roughness of the lenses deposited on a SiO/Si substrate is determined by an atomic force microscope (AFM). The round surface exhibits a typical deviation from a spherical surface of less than 3% with a surface roughness of <1 nm (Fig. 1d). The nearly perfect PSC structure demonstrates that self-assembled CHQ lenses are high-quality optical elements.

Figure 1e shows an optical microscope image of a CHQ lens ($D = 970$ nm, $H = 220$ nm) on top of CHQ tubule bundles under filtered light ($\lambda_{\max} = 472$ nm) from a halogen lamp. The image indicates that the lens magnifies the underlying object with a magnification factor $M (=l'/l = \sim 1.6)$. The paraxial focal length estimated from the observed magnification is $F = HM/(M - 1) = 590$ nm, much shorter than that expected from geometric optics, $F_{\text{geo}} = R/(n - 1) = 1.3 \mu\text{m}$ (R , lens radius; $n = 1.5$, refractive index). The deviations from geometrical optics for the subwavelength-size CHQ lenses are the key signature of near-field focusing. The reduction of the focal length in the CHQ nanolens implies enhanced magnifying effects through the nanolens.

To demonstrate enhanced spatial resolution, we further investigate the optical properties of CHQ lenses on pre-fabricated

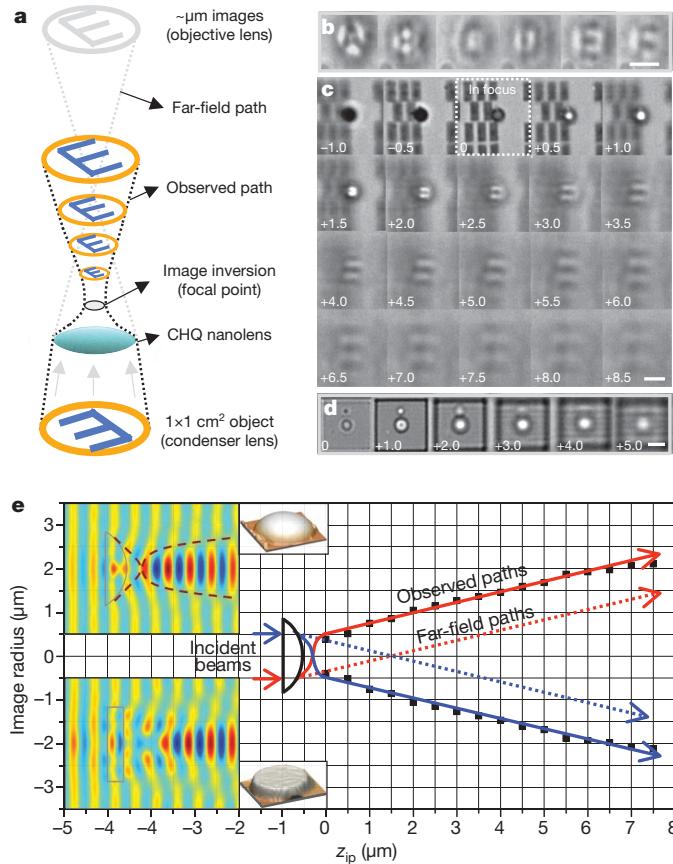


Figure 3 | Optical images and beam trajectories of alphabetical characters projected through CHQ lenses and PMMA disks. **a**, Schematic illustration of the imaging process. The light sources transmitting a patterned glass are collimated by a condenser lens. As the nanolens focuses the beam, inverted images are formed after passing through the focal point. Each slice of the images is obtained at a distance z_{ip} from the lens/disk bottom. **b**, Optical microscope images of various inverted alphabetical characters (A–F) projected through the lens. The upright non-inverted characters are imaged due to the image inversion at the image plane $\sim 3.5 \mu\text{m}$ away from the nanolens. **c**, Optical microscope images of the alphabetical character "E" projected through a CHQ lens ($D = 1.7 \mu\text{m}$, $H = 480$ nm) on vertical image planes at varying distances z_{ip} . The number in each slice denotes the value of z_{ip} . "In-focus" denotes the image focused on the plane of the Pd stripe patterns. **d**, Projected optical microscope images through a $1.7\text{-}\mu\text{m}$ -wide and 400-nm-thick PMMA disk, which shows no clear alphabetical character image formation. **e**, Beam trajectory with reduced focal length in the near-field PSC lens. Small insets on the left, AFM images of the CHQ lens (upper) and the PMMA disk (lower). Large insets on the left, FDTD simulation results of the radial component of the electric field (E_x) of the PSC lens (upper) and the PMMA disk (lower) ($\lambda = 472$ nm). All scale bars, 2 μm.

subwavelength objects as follows (Fig. 2). First, 250- and 220-nm pitch metallic (Pd/Cr, 120/3 nm thick) stripe arrays are fabricated by electron beam lithography on glass substrates. The CHQ lenses isolated in aqueous suspensions (Fig. 1c) are randomly placed on the patterned substrate by spinning (Fig. 2b, d, f). We use a high resolution optical microscope to take reflection mode images of the same part of the sample (Fig. 2a, c, e). The optical images were obtained through a $100\times$ objective lens with a numerical aperture (NA) of 0.9. The Rayleigh resolution limit for point objects ($r = 0.61\lambda/\text{NA}$) is 320 nm, while that for line objects ($r = 0.5\lambda/\text{NA}$) is 262 nm. A more stringent Sparrow resolution limit ($r = 0.475\lambda/\text{NA}$)²⁰ is 249 nm. Indeed, the optical images outside the CHQ lenses do not resolve the underlying metallic stripes, as the stripe spacings of $d = 220$ and 250 nm are narrower than, or very similar to, the stringent resolution limit. On the other hand, resolved individual metallic stripes are clearly imaged through the CHQ lens (Fig. 2a, c for 250 nm and Fig. 2e for 220 nm). The image magnification increases as the distance

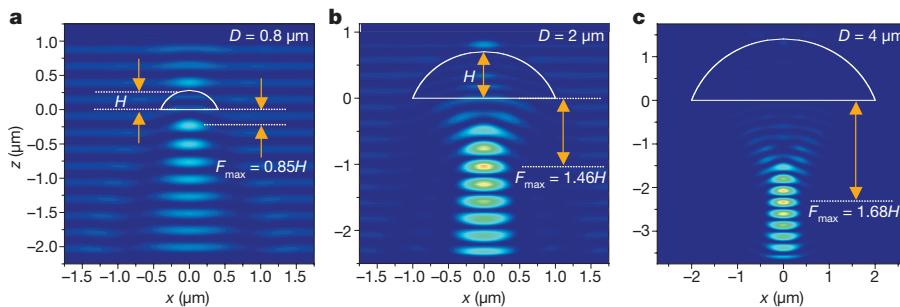


Figure 4 | Focal length changes for various sizes of CHQ lenses (fixed $H/D = 0.35$). a, $D = 0.8 \mu\text{m}$; b, $D = 2 \mu\text{m}$; c, $D = 4 \mu\text{m}$. Data were obtained from FDTD simulation results of $|E_x|^2$ ($\lambda = 472 \text{ nm}$).

between nanolens and image is increased with micromanipulators and piezo-controlled nano-positioning stages. The magnifying effect enhances the resolution substantially (by as much as 2.5 times). The magnified images of the face-up lenses show pin-cushion distortion (Supplementary Information C), whereas no notable distortion appears for the face-down lens. This difference arises because the near-field image of the face-up lens is formed by the interference of the secondary Fresnel's waves on the flat and convex surfaces of the lens, while the near-field image of the face-down lens is formed by the secondary surface waves due to the convex surface.

To analyse the imaging and focusing behaviours through the CHQ lenses, we projected a series of alphabetical character images into a CHQ lens ($D = 1.7 \mu\text{m}$, $H = 0.48 \mu\text{m}$) via a condenser lens ($\text{NA} = 0.8$) in the far field (Fig. 3a–c). The images of “E” formed by the PSC lens (Fig. 3c) are compared with those formed by a flat disk (Fig. 3d). The transmitted images through the PSC lens or flat disk were recorded by an optical microscope with a CCD camera focused on different image planes at a distance z_{ip} from the lens/disk bottom (Supplementary Information F). In Fig. 3e, the solid lines are guides to the eye, following the square dots (which are the measured optical beam trajectories of the top and bottom edges of the “E” image along z_{ip}). The far-field optical paths calculated from geometrical ray optics (dotted lines) are also presented for comparison. As the image plane moves away from the CHQ lens, the images are inverted and magnified. In Fig. 3c, we observe clear magnified and inverted images of “E” along the axial positions for $z_{\text{ip}} \geq 1.5\text{--}2.0 \mu\text{m}$, and the letter shape is still seen at $z_{\text{ip}} = 1.0 \mu\text{m}$. At $z_{\text{ip}} < 0.5 \mu\text{m}$, we observe that the bright spot with a dark annular ring at $z_{\text{ip}} = 0.5 \mu\text{m}$ changes to the grey/dark spot with a bright annular ring surrounding the dark spot. We did not observe the direct non-inverted image for $z_{\text{ip}} < 0.5 \mu\text{m}$. However, since the “E” image is clearly inverted for $z_{\text{ip}} \geq 1.5 \mu\text{m}$, we deduce that the image inversion should occur at the focal point between the two adjacent points at $z_{\text{ip}} = 0.5 \mu\text{m}$ (lens top) and $z_{\text{ip}} = 1.0 \mu\text{m}$, which have the first and second smallest bright spot radii (0.38 and $0.42 \mu\text{m}$, respectively) among many images taken at every $0.5\text{-}\mu\text{m}$ step.

As the lens height is $0.48 \mu\text{m}$ and the bright spot size at $z_{\text{ip}} = 0.5 \mu\text{m}$ is slightly smaller than that at $z_{\text{ip}} = 1.0 \mu\text{m}$, the focal point is expected to be located at $0.48 \mu\text{m} < z_{\text{ip}} < 0.75 \mu\text{m}$ (then the focal length is not more than $0.27 \mu\text{m}$). Thus, this focal length is in agreement with the theoretically estimated focal length ($0.3 \mu\text{m}$, or $z_{\text{ip}} = \sim 0.8 \mu\text{m}$) assumed as the first Fresnel zone focal length ($F = 4\lambda/(\pi\text{NA})^2$). Given the axial resolution limit (Rayleigh range $z_R = (4/\pi)\lambda/\text{NA}^2$) of $0.9 \mu\text{m}$ according to Gaussian beam optics, the measured focal point is at $z_{\text{ip}} = \sim 0.7 \pm 0.9 \mu\text{m}$, or $F = \sim 0.2 \pm 0.9 \mu\text{m}$, much smaller than the geometrical focal length F_{geo} ($2.0 \mu\text{m}$; $z_{\text{ip}} = \sim 2.5 \mu\text{m}$). This drastically shortened focal length ($F \ll F_{\text{geo}}$), which should yield the curvilinear trajectories for $z_{\text{ip}} < 1.0 \mu\text{m}$, is supported in Fig. 3e by tracing the image size of the “E” in Fig. 3c down to $z_{\text{ip}} = 1.0 \mu\text{m}$. It is further clearly seen in accurate finite-difference time domain (FDTD) simulations that the inversion occurs at $z_{\text{ip}} = 0.8 \mu\text{m}$ (Fig. 3e top left inset). Here, the images between $z_{\text{ip}} = 0.0 \mu\text{m}$ and

the focal point are in the Fresnel diffraction region, appearing as the Fourier transformed image. In this region, the asymptotic beam path is curvilinear so that the image can be inverted at the short focal point. This result explains the experiment showing the inverted image of “E” consistently obtained for $z_{\text{ip}} \geq 1.0 \mu\text{m}$ in Fig. 3c.

Shape-dependent focusing and imaging was studied by performing a similar experiment through a patterned poly(methylmethacrylate) (PMMA) disk formed by electron beam lithography. The disk has flat surfaces on both sides, but its thickness and size are similar to those of the CHQ lens. In contrast to the clear images created through the PSC lens, the light propagating through the disk shows no clear images (Fig. 3d). These experimental images are consistent with the electromagnetic simulation (Fig. 3e bottom left inset); at $z_{\text{ip}} = \sim 0 \mu\text{m}$ along the axis, the image has a small bright spot; at $z_{\text{ip}} = \sim 1 \mu\text{m}$, the image has a small dark spot at the centre and an annular bright spot; and the images for $z_{\text{ip}} \geq \sim 2 \mu\text{m}$ again have bright spots at the centre. The difference between the images created through the PSC lens and the flat disk is thus clearly confirmed by electromagnetic simulations, and the image formation depends on the shape (surface curvature) of the optical elements. We note that the focal point of a nanolens originates mainly from surface waves at the spherical interface of the PSC lens, and the near-field focal length is drastically shortened by the interference of the propagating waves from the lens edges because the nanolens length-scale is comparable to the wavelength (Supplementary Information B). The resulting curvilinear ray path with a wave propagation to match the amplitude and phase inside and outside the lens forms a small focused spot at a very short near-field focal distance (solid lines in Fig. 3e).

To investigate the size-dependent diffraction/refraction phenomena in the near-field regime, we performed FDTD simulations²¹ for sub-wavelength-size lenses of different diameters ($D = 0.8\text{--}4 \mu\text{m}$) with fixed ratio of $H/D = 0.35$ (Fig. 4). The incident plane waves are polarized along the x -direction, and Fig. 4 shows the spatial distribution of $|E_x|^2$ where E_x is the x -component of the electric field. As the lens size approaches the wavelength, the near-field focal length (F_{max}) showing the maximum light intensity decreases remarkably towards the first Fresnel zone. This near-field focusing is due to superposition of the diffraction on the flat aperture and the interference of secondary Fresnel waves on the spherical surface of the PSC lens, whose diameter is comparable to the wavelength of light.

Given that high resolution imaging beyond or near the diffraction limit has been achieved by relying on a stimulated-emission-depletion method, a lens-less near-field optical method, and novel materials approaches^{22–26}, the near-field focusing and magnification discussed here represent a complementary approach to obtaining lens-based high resolution beyond the diffraction limit at low intensities. This magnification can be further increased by using immersion lensing techniques. Such a combination would lead to hyper-refraction phenomena due to surface waves on the interfaces. Based on the merit of spherical lens-based optics, the near-field focusing and magnification phenomena in nanoscale lenses would have wide applications, including super-resolution by a nanolens array and by a nanolens on an AFM

or aperture-less near-field scanning optical microscope (ANSOM). This is possible because nanolenses at arbitrary locations can successfully be transferred to pre-determined locations by using micromanipulators with the assistance of an SEM, and fabricated in array in order to increase the area over which high spatial resolution is achieved (Supplementary Information E). In particular, the focusing and magnification effect of the nanolens would have useful applications in signal enhancements in spectroscopy—in micro-photoluminescence intensity with quantum dots for single (near-infrared) photon spectroscopy (Supplementary Information H), and in Raman intensity on substrates such as graphene^{27,28} (Supplementary Information I). Application to deep ultraviolet lithography is also possible (Supplementary Information G), because regardless of the visible or ultraviolet wavelength, nanolenses can be used to obtain high resolution beyond the diffraction limit, as long as the wavelength is comparable to the lens size.

METHODS SUMMARY

CHQ nanoscale lenses are self-assembled in 1:1 acetone–water solution by evaporating the solution slowly for a few days in ambient conditions. After dissolving 10 mg CHQ monomers in 2 ml acetone solvent, 2 ml water is added. Slow evaporation of solution leads to the self-assemblies of CHQ nanostructures including CHQ lenses. Pd/Cr (120/3 nm thick, 220 and 250 nm pitch) line patterns are fabricated on a 0.13–0.16 mm thick glass substrate with e-beam lithography. After depositing chromium to a thickness of 3 nm using a thermal evaporator (BOC Auto 306, Edwards), positive electron-beam resist PMMA is spin-coated on a SiO₂ substrate at 4,000 r.p.m. The PMMA film is exposed to 30 kV electron beam and developed with MIBK:IPA = 1:3 developer for patterning. Deposition of 120 nm palladium is performed with an electron-beam evaporator (SC2000, SEMICORE), and then the patterns are accomplished through a lift-off process. We obtained the magnified image through the CHQ lens using an optical microscope equipped with micromanipulators and piezo-controlled nano-positioners (E-610, Physik Instrumente). FDTD simulations were performed using the FullWAVE 4.0 program (RSoft Design Group). In the Supplementary Information we describe relocation of CHQ lenses with a dual focused ion beam microscope (Helios Nanolab 600, FEI) and an SEM (JSM6390, Jeol). CHQ lenses are attached to the end of a tungsten tip (which is adhesive after contact with an adhesive carbon tape), and then transferred to pre-determined positions. The enhancement of Raman intensity of graphene through a CHQ lens is observed by a micro-Raman microscope (InVia Raman microscope, Renishaw; laser wavelength 633 nm, power 4 mW).

Received 30 December 2008; accepted 28 May 2009.

1. Smolyaninov, I. I., Hung, Y.-J. & Davis, C. C. Magnifying superlens in the visible frequency range. *Science* **315**, 1699–1701 (2007).
2. Liu, Z., Lee, H., Xiong, Y., Sun, C. & Zhang, X. Far-field optical hyperlens magnifying sub-diffraction-limited objects. *Science* **315**, 1686 (2007).
3. Fang, N., Lee, H., Sun, C. & Zhang, X. Sub-diffraction-limited optical imaging with a silver superlens. *Science* **308**, 534–537 (2005).
4. Hell, S. W. Far-field optical nanoscopy. *Science* **316**, 1153–1158 (2007).
5. Yano, T., Shibata, S. & Kishi, T. Fabrication of micrometer-size glass solid immersion lens. *Appl. Phys. B* **83**, 167–170 (2006).
6. Merlin, R. Radiationless electromagnetic interference: evanescent-field lenses and perfect focusing. *Science* **317**, 927–929 (2007).
7. Whitesides, G. M. & Grzybowski, B. Self-assembly at all scales. *Science* **295**, 2418–2421 (2002).
8. Aizenberg, J., Tkachenko, A., Weiner, S., Addadi, L. & Hendler, G. Calcitic microlenses as part of the photoreceptor system in brittlestars. *Nature* **412**, 819–822 (2001).
9. Lee, L. P. & Szema, R. Inspirations from biological optics for advanced photonic systems. *Science* **310**, 1148–1150 (2005).
10. Dong, L., Agarwal, A. K., Beebe, D. J. & Jiang, H. Adaptive liquid microlenses activated by stimuli-responsive hydrogels. *Nature* **442**, 551–554 (2006).

11. Yang, S. K. U., Chaitanya, T. E. L., Chen, G. & Aizenberg, J. Microlens arrays with integrated pores as a multipattern photomask. *Appl. Phys. Lett.* **86**, 201121 (2005).
12. Jeong, K.-H., Kim, J. & Lee, L. P. Biologically inspired artificial compound eyes. *Science* **312**, 557–561 (2006).
13. Fletcher, D. A., Goodson, K. E. & Kino, G. S. Focusing in microlenses close to a wavelength in diameter. *Opt. Lett.* **26**, 399–401 (2001).
14. Wu, H., Odom, T. W. & Whitesides, G. M. Connectivity of features in microlens array reduction photolithography: generation of various patterns with a single photomask. *J. Am. Chem. Soc.* **124**, 7288–7289 (2002).
15. Hong, B. H. et al. Self-assembled arrays of organic nanotubes with infinitely long one-dimensional H-bond chains. *J. Am. Chem. Soc.* **123**, 10748–10749 (2001).
16. Kim, K. S. et al. Assembling phenomena of calix[4]hydroquinone nanotube bundles by one-dimensional short hydrogen bonding and displaced π-π stacking. *J. Am. Chem. Soc.* **124**, 14268–14279 (2002).
17. Kim, K. S., Tarakeshwar, P. & Lee, J. Y. Molecular clusters of π-systems: theoretical studies of structures, spectra and origin of interaction energies. *Chem. Rev.* **100**, 4145–4186 (2000).
18. Hong, B. H., Bae, S. C., Lee, C.-W., Jeong, S. & Kim, K. S. Ultrathin single-crystalline silver nanowire arrays formed in an ambient solution phase. *Science* **294**, 348–351 (2001).
19. Fujita, J., Ohnishi, Y., Ochiai, Y. & Matsui, S. Ultrahigh resolution of calixarene negative resist in electron beam lithography. *Appl. Phys. Lett.* **68**, 1297–1299 (1996).
20. Goldstein, D. J. Resolution in light microscopy studied by computer simulations. *J. Microsc.* **166**, 185–197 (1992).
21. Taflove, A. & Hagness, S. C. *Computational Electrodynamics: The Finite-difference Time-domain Method* (Artech House, 2000).
22. Simpson, G. J. Biological imaging: the diffraction barrier broken. *Nature* **440**, 879–880 (2006).
23. Lewis, A. et al. Near-field optics: from subwavelength illumination to nanometric shadowing. *Nature Biotechnol.* **21**, 1378–1386 (2003).
24. Smith, D. R., Pendry, J. B. & Wilshire, M. C. K. Metamaterials and negative refractive index. *Science* **305**, 788–792 (2004).
25. Parimi, P. V., Lu, W. T., Vodo, P. & Sridhar, S. Photonic crystals: imaging by flat lens using negative refraction. *Nature* **426**, 404 (2003).
26. Chatterjee, R. et al. Achieving subdiffraction imaging through bound surface states in negative refraction photonic crystals in the near-infrared range. *Phys. Rev. Lett.* **100**, 187401 (2008).
27. Novoselov, K. S. et al. Electric field effect in atomically thin carbon films. *Science* **306**, 666–669 (2004).
28. Kim, K. S. et al. Large-scale pattern growth of graphene films for stretchable transparent electrodes. *Nature* **457**, 706–710 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank T. F. Heinz, C. K. Hong, J. H. Lee and W. J. Kim for discussions, and K. Cho, J. T. Han, J. W. Lee and C. S. Lee for assisting in characterization. This work was supported by the Korea Foundation for International Cooperation of Science and Technology (Global Research Laboratory programme), Korea Science and Engineering Foundation grants funded by the Korea Government (World Class University, R32-2008-000-10180-0, R33-2008-000-10138-0; EPB Center, 2009-0063312; 2009-0062808; 2009-0060271), the Brain Korea 21 (Korea Research Foundation), the National Science Foundation (NSF: CHE-0641523; ECCS-0747787) and the New York State Office of Science (NYSTAR).

Author Contributions J.Y.L. and B.H.H. conducted experiments (synthesis, characterization, optical measurements). Y.K. assisted in synthesis. R.B., B.H.H. and C.W.W. conducted electromagnetic simulations, and W.Y.K., S.K.M. and M.V.J. analysed the simulation results. L.J.K. assisted in the high-resolution optical imaging analysis. I.-C.H. conducted lens transfer and lens array formation. Keun S. Kim and J.Y.L. obtained micro-Raman spectra. P.K. supervised optical measurements. Kwang S. Kim supervised the whole project.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to Kwang S. Kim (kim@postech.ac.kr) or P.K. (pkim@phys.columbia.edu).

LETTERS

Formation of evenly spaced ridges and valleysJ. Taylor Perron¹, James W. Kirchner^{2,3,4} & William E. Dietrich²

One of the most striking examples of self-organization in landscapes is the emergence of evenly spaced ridges and valleys^{1–6}. Despite the prevalence of uniform valley spacing, no theory has been shown to predict this fundamental topographic wavelength. Models of long-term landscape evolution can produce landforms that look realistic^{7–9}, but few metrics exist to assess the similarity between models and natural landscapes. Here we show that the ridge–valley wavelength can be predicted from erosional mechanics. From equations of mass conservation and sediment transport, we derive a characteristic length scale at which the timescales for erosion by diffusive soil creep and advective stream incision are equal. This length scale is directly proportional to the valley spacing that emerges in a numerical model of landform evolution, and to the measured valley spacing at five field sites. Our results provide a quantitative explanation for one of the most widely observed characteristics of landscapes. The findings also imply that valley spacing is a fundamental topographic signature that records how material properties and climate regulate erosional processes.

The spacing between adjacent ridges and valleys is a fundamental dimension of hilly topography^{1–6}. Even a casual observer can see from an aeroplane window that ridges and valleys in many landscapes appear to be uniformly spaced (Fig. 1), even where their locations are not controlled by bedrock structure. Indeed, uniform spacing is often most clearly visible where bedrock is mechanically homogeneous¹⁰. This implies that the characteristic ridge–valley wavelength is an emergent property of the erosion and sediment transport processes that shape the landscape. Any theory for the long-term evolution of Earth's surface should be able to explain fundamental landscape scales like the ridge–valley wavelength.

Some of the earliest theories of landscape evolution focused on the segmentation of landscapes into ridges (or, more generally, hillslopes) and valleys. Davis¹¹ and Gilbert^{10,12} suggested that hillslopes are dominated by sediment transport mechanisms that smooth the land surface, and that hillslopes give way to valleys where water runoff becomes concentrated enough to outpace the smoothing processes and incise into the land surface. Later studies showed how this competition might lead to the development of evenly spaced valleys. Smith and Bretherton¹³ demonstrated that a concave-up, erodible surface under a sheet of flowing water—a situation analogous to a freshly exposed soil embankment during a rainstorm—is unstable with respect to perturbations, with the shortest-wavelength topographic features growing fastest. This result implied no preferred wavelength. Subsequent studies found that if smoothing is introduced, either by diffusive processes^{14,15} such as rain splash¹⁶ or by the dispersive effects of the free water surface², perturbations with an intermediate wavelength will grow fastest, forming incipient erosional rills with a characteristic spacing. Some studies additionally included a sediment transport threshold that encouraged the selection of an intermediate wavelength^{2,5}.

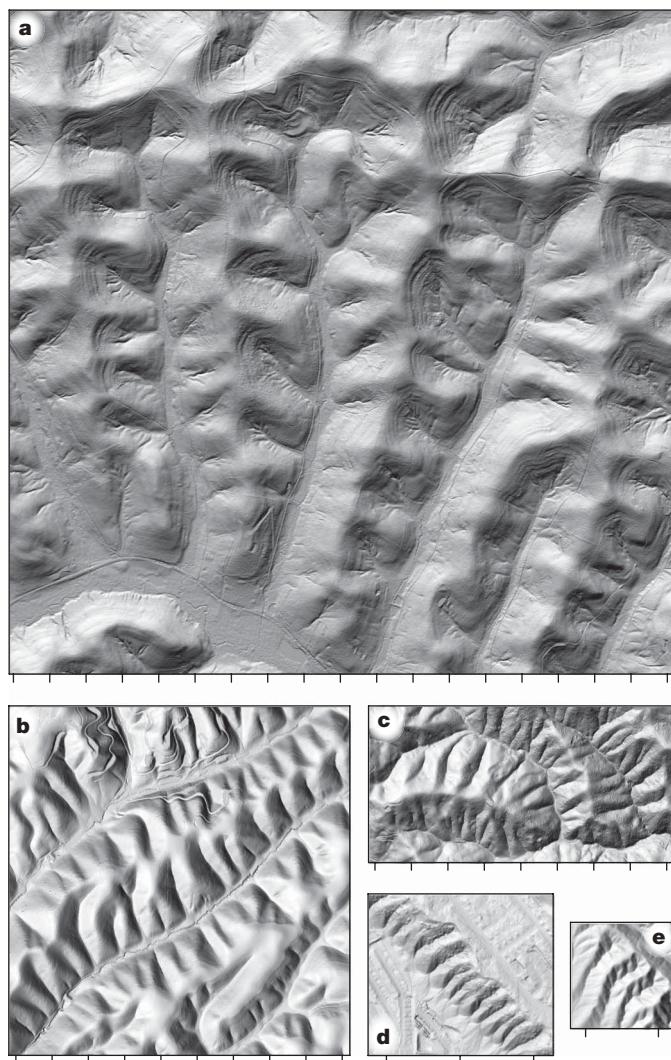


Figure 1 | Uniform valley spacing. Shaded relief maps of representative sections of the study sites. **a**, Eaton Hollow, Pennsylvania. **b**, Gabilan Mesa, California. **c**, Napa Valley, California. **d**, Point of the Mountain, Utah. **e**, Dragon's Back, California. Tick spacing is 200 m. For clarity, **d** and **e** have been enlarged by a factor of two relative to **a–c**. Vegetation has been filtered out of the data to reveal the underlying topography. Eaton Hollow data are from the State of Pennsylvania PAMAP program; Point of the Mountain data are from the State of Utah Automated Geographic Reference Center; California data are from the National Center for Airborne Laser Mapping (NCALM).

¹Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ²Department of Earth and Planetary Science, University of California, Berkeley, California 94720, USA. ³Swiss Federal Institute for Forest, Snow and Landscape Research (WSL), Birmensdorf, Switzerland. ⁴Department of Environmental Sciences, Swiss Federal Institute of Technology (ETH), Zürich, Switzerland.

These studies considered incipient channelization of a surface by a sheet of flowing water, and cannot be used to predict the dimensions of large-scale landforms like those in Fig. 1. Numerical models based on a similar competition between stream channel incision and diffusive soil creep have been used to explore the long-term evolution of such landforms, including the factors controlling the upslope drainage area at which hillslopes transition into valleys^{8,9,17–21}, but not the characteristic ridge-valley wavelength. Moreover, comparisons between models and natural landscapes have been hampered by the scarcity of high-resolution topographic data and the difficulty of measuring the long-term rates of erosion and transport processes in the field.

To investigate the factors that control valley spacing, we developed a numerical model (Methods, Supplementary Information) that simulates landscape evolution under the combined influence of soil creep (here used to mean downslope soil flux due to abiotic and biotic processes, linearly dependent on the local surface gradient) and stream incision. The transient evolution of the model illustrates how uniform valley spacing emerges over time (Supplementary Information). As the topography evolves from a randomly rough, horizontal initial surface, irregularly spaced incipient valleys form at the boundaries and begin to grow by lengthening and widening. Competition for drainage area (a proxy for water flux) stunts the growth of valleys that are too small or spaced too closely together. This transient evolution is similar in many respects to early conceptual models of drainage network development^{10,16,22}. The model eventually reaches a deterministic equilibrium in which the spacing of valleys is approximately uniform.

Nonlinearities in the governing equation (equation (1)) preclude an analytic solution for the equilibrium topography, so we used dimensional analysis to explore how the erosion and transport parameters control the valley spacing. The governing equation is a nonlinear advection-diffusion equation, and we derived a quantity analogous to a Péclet number, Pe (equation (2)), that expresses the relative magnitudes of the advective stream incision and diffusive soil creep mechanisms shaping the landscape. When Pe is small, the landscape is dominated by creep, and forms a smooth slope with no valleys. When Pe is large, the landscape is dominated by stream incision, and forms networks of branching valleys. Setting $Pe = 1$ yields a characteristic length scale, L_c (equation (3)), at which the characteristic timescales for stream incision and creep are equal. Numerical modelling has shown that L_c^2 is approximately the drainage area at which the topography transitions from a concave-down, creep-dominated hillslope to a concave-up, stream-incision-dominated valley²³. We computed equilibrium model solutions using parameters that give a range of values for L_c , and measured valley spacing, λ , by identifying the dominant peaks in the two-dimensional Fourier spectra of the simulated topography^{6,23}. For a given value of L_c , a range of valley spacing is possible because a range of slope lengths can give rise to first-order valleys (valleys that do not branch), and longer slopes form more widely spaced valleys (Fig. 2). The range of slope lengths is limited, however, because slopes that are too long will become dissected by branching valleys, and slopes that are too short will remain smooth and undissected. The minimum and maximum valley spacings are directly proportional to L_c , as shown in Fig. 2.

To test whether this theoretical prediction is consistent with valley spacing in natural landscapes, we measured L_c in five landscapes in the United States that have different valley spacings: Gabilan Mesa and Napa Valley in the California Coast Ranges, the Dragon's Back pressure ridge along the San Andreas fault in the Carrizo Plain, California, Point of the Mountain in Salt Lake Valley, Utah, and Eaton Hollow in southwestern Pennsylvania. All five sites display uniform valley spacing (Fig. 1, Table 1) that is not determined by structural heterogeneities in the underlying bedrock. Gabilan Mesa is an oak savannah with a Mediterranean climate, and erosion of the moderately consolidated sandstones, siltstones and conglomerates of the Paso Robles Formation has formed valleys with a spacing of

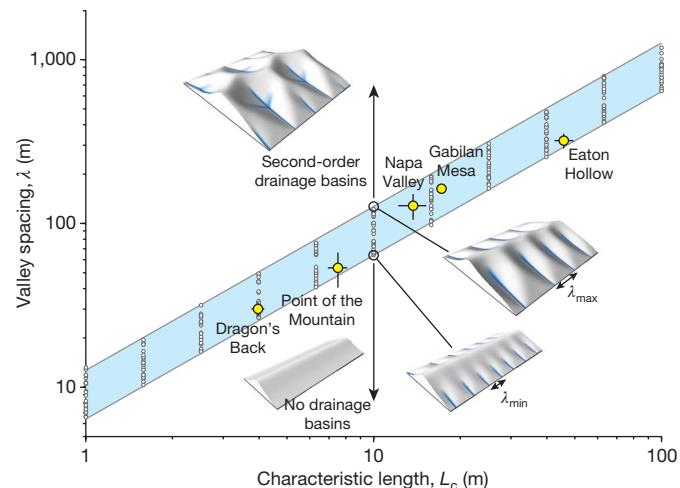


Figure 2 | Comparison of predicted and observed valley spacing. Plot of valley spacing, λ , against the characteristic length scale, L_c (equation (3) in Methods), for first-order drainage basins. Each white circle represents the valley spacing in a single numerical model solution. The blue trend shows the range of possible valley spacings, which correspond to different slope lengths, for each value of L_c . Slope length was controlled by varying the width of the model grid in the direction normal to the main ridgeline. The minimum and maximum spacing for a given value of L_c correspond to the shortest and longest slopes that form first-order valleys. The expression for the blue trend is $6.4L_c \leq \lambda \leq 12.7L_c$. Insets are perspective views of numerical model solutions with the same L_c but different slope lengths, with valley bottoms shaded blue. Yellow points are the means for first-order valleys in the study sites. Error bars are one standard error of the mean.

163 ± 11 m. Napa Valley has similar vegetation and climate, with valleys spaced at 128 ± 23 m that have formed in sandstones and mudstones of the Franciscan Complex. Dragon's Back is a semi-arid grassland underlain by sediments of the Paso Robles Formation that are less consolidated than in Gabilan Mesa, with a valley spacing of 30 ± 3 m. Point of the Mountain is a sand and gravel spit formed by Pleistocene Lake Bonneville, with a valley spacing of 54 ± 13 m. Eaton Hollow is a temperate mixed forest underlain by horizontal beds of Permian and Pennsylvanian sandstone, shale, limestone and coal, with a valley spacing of 321 ± 33 m. Valley spacings were measured from peaks in two-dimensional Fourier spectra derived from high-resolution laser altimetry maps⁶. Comparison with spectra for random surfaces with the same roughness characteristics as the observed topography⁶ shows that valley spacing as uniform as that observed in the study sites is very unlikely to arise by chance ($P < 0.001$).

Erosion and transport at all five sites are dominated by stream channel incision and by diffusive soil creep, which occurs mainly through bioturbation such as tree throw and rodent burrowing. Mean hillslope gradients are between 0.2 and 0.4, and evidence of landslides is rare. In Napa Valley, some of the areas surrounding our study site are steeper and have experienced landslides, but we avoided these areas in our analysis. Similarly, portions of Dragon's Back experience nonlinear creep and frequent landslides due to a spatial gradient in uplift rates²⁴, but we restricted our analysis to the drainage basins farthest from the zone of maximum uplift, which are dominated by linear creep. The mechanically homogeneous substrates and the two dominant erosion and transport mechanisms conform to the simplifying assumptions behind the numerical model, making these sites suitable locations to test predictions of valley spacing.

An estimate of L_c for each landscape requires values for the constants that describe the long-term strengths of the erosional processes: soil diffusivity D , stream erosivity K , and drainage area exponent m (equation (3)). These parameters are difficult to measure directly because erosion is usually slow or episodic, and because present-day rates may not be representative of long-term rates. We

Table 1 | Topographic measurements

	$\nabla^2 z_h (10^{-3} \text{ m}^{-1})$	$D/K (\text{m}^{2m+1})$	m	$L_c (\text{m})$	$\lambda (\text{m})$	Mean annual precipitation* (mm)
Dragon's Back	-94 ± 3	12 ± 1	0.42 ± 0.01	4.0 ± 0.2	30 ± 3	0.23
Point of the Mountain	-28 ± 4	26 ± 3	0.31 ± 0.02	7.5 ± 0.6	54 ± 13	0.50
Napa Valley	-18.8 ± 0.3	86 ± 13	0.35 ± 0.02	13.7 ± 1.5	128 ± 23	0.93
Gabilan Mesa	-11.8 ± 0.4	124 ± 3	0.35 ± 0.003	17.2 ± 0.4	163 ± 11	0.32
Eaton Hollow	-5.5 ± 0.1	802 ± 82	0.37 ± 0.01	46.0 ± 3.4	321 ± 33	1.05

*For the period 1971–2000. Data are from the PRISM Climate Group, Oregon State University, <http://prismclimate.org>.

therefore used the shapes of hilltops and stream profiles measured from high-resolution topography to solve for time-averaged values of D/K and m (Fig. 3, Methods).

The comparison in Fig. 2 shows that valley spacing is proportional to L_c across all five study sites, consistent with the predictions of the numerical model. This good agreement suggests that the two-process model, despite its simplifications, captures the mechanisms that exert the strongest influence on valley spacing in these landscapes. To demonstrate that this agreement is not an inevitable consequence

of our procedure for measuring L_c , we performed the same topographic measurements in three landscapes shaped by erosional processes that are not well described by our model; the valley spacing in those landscapes is inconsistent with the inferred values of L_c (Supplementary Information).

Our measurements and the geology and climate of the study sites offer some insight into the differences in L_c and valley spacing. In our model, longer L_c and wider valley spacing can result from larger D , smaller K , or smaller m (equation (3)). Our topographic measurements (Table 1) indicate that the drainage area exponent m is similar for the five sites, and that the differences in valley spacing primarily reflect differences in D/K , the ratio of soil diffusivity to stream erosivity. Systematic variations in bedrock mechanical strength among the five sites further suggest that rock strength, which we expect to be negatively correlated with K , is a major source of variability in D/K . Sites with the least consolidated sediments (Dragon's Back and Point of the Mountain) have the narrowest valley spacing, sites with moderately consolidated sediments (Gabilan Mesa and Napa Valley) have intermediate spacing, and the site with the most competent bedrock (Eaton Hollow) has the widest valley spacing.

A comparison of precipitation rates at the five sites (Table 1) suggests that climate may also influence valley spacing: with the exception of Gabilan Mesa, wider valley spacing corresponds to greater present-day mean annual precipitation. One possible cause is the stream erosivity, which depends on drainage basin hydrology as well as on rock strength^{7–9,23}. Although higher rainfall should increase streamflow, the dominant effect of precipitation in soil-mantled landscapes like those analysed here may be to reduce K by promoting vegetation growth and infiltration, thereby inhibiting overland flow erosion⁹. It also seems likely that more intense bioturbation in wetter environments leads to higher soil diffusivity⁹, an effect consistent with previous measurements²⁵ of D and with the observed correlation between precipitation and the hilltop curvature, $\nabla^2 z_h$ (Table 1, Methods). Although we are at present unable to quantify the relative importance of these mechanisms, our observations suggest that valley spacing may serve as a topographic proxy for the combined effects of bedrock mechanical strength and climate on the relative magnitudes of different erosional processes.

It is notable that our theory closely predicts valley spacing in the five study sites even though it does not include a threshold for fluvial erosion. Soil cohesion and plant roots impart strength to the soil surface, such that very small flows may not exert enough stress to erode the underlying material. There is evidence that such thresholds influence the locations of fluvial channel heads²⁶, and it has been proposed that thresholds also influence the scale at which hillslopes transition into valleys^{22,27–29}. If a fluvial erosion threshold is included in the model equations, its effect is generally to widen valley spacing²³, though not as much as a comparable fractional change in D or K . The fact that our model does not systematically underpredict valley spacing suggests that competition between soil creep and stream incision is the primary mechanism that controls valley spacing in these landscapes, but we acknowledge that erosion thresholds could have a stronger influence in others.

Also notable is the prediction that L_c , and therefore valley spacing, is independent of erosion rate. This is consistent with previous observations^{8,9,23} that steady-state drainage density is independent of erosion rate when both creep flux and stream incision rate vary

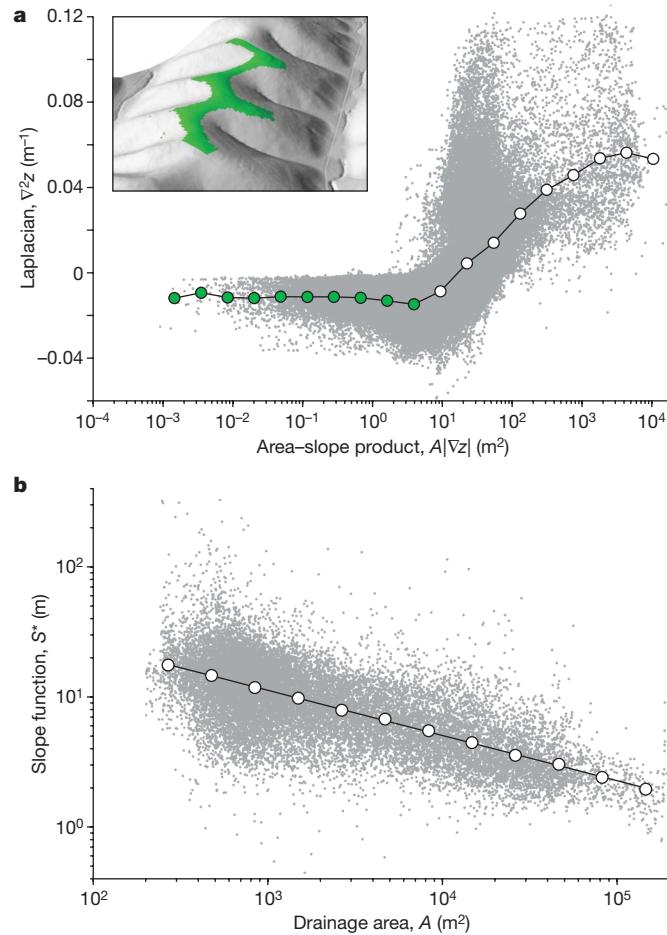


Figure 3 | Measurement of model parameters from topography. **a**, Plot of the Laplacian of elevation against the product of drainage area and slope for first-order drainage basins in Gabilan Mesa. Filled circles are means of log-transformed data within logarithmically spaced bins. On hilltops, where both drainage area and slope are small (green shading), the Laplacian is roughly constant, consistent with equilibrium topography (equation (4), Methods). The inset shows several representative hilltops. For clarity, the plot shows a random subsample of 25% of the raw data points. **b**, Plot of slope function (equation (5), Methods) against drainage area for stream profiles in the same basins. Filled circles are means of log-transformed data within logarithmically spaced bins, and the line is a least-squares fit to the binned data. Error bars showing one standard error of the mean for the filled circles are smaller than the symbols. Plots for all study sites are shown in the Supplementary Information.

linearly with topographic slope (equation (1)). The trend in Fig. 2, defined by sites that probably have different erosion rates, suggests that this linearity assumption is reasonable.

Valley spacing is a fundamental topographic signature that varies widely across the Earth^{1–6,23} and other planetary surfaces²³. The simplified yet mechanistic approach introduced here enables us to predict valley spacing by parameterizing erosion and transport expressions through topographic analysis. This analysis shows that differences in valley spacing are linked to elementary ratios of rate coefficients that may in turn depend on geologic materials and climate, two regulators of landform evolution that are currently poorly quantified in erosion theory. Thus, valley spacing is a measurable clue to aspects of a site's geologic past that can otherwise be difficult to assess. The same may be true of other emergent patterns in landscapes.

METHODS SUMMARY

Numerical model. Following several previous studies^{7–9,17}, we describe the evolution of the topography with a nonlinear advection–diffusion equation:

$$\frac{\partial z}{\partial t} = D\nabla^2 z - KA^m |\nabla z| + U \quad (1)$$

where z is elevation, t is time, D is soil diffusivity, A is drainage area, K and m are constants, and U is surface uplift rate. Equation (1) assumes that soil creep flux is proportional to the local topographic gradient, and that the rate of erosion by channelized flow of water is proportional to the rate of energy expenditure³⁰. A derivation of equation (1) and details of the numerical solution method can be found in ref. 23. Non-dimensionalizing equation (1) yields a quantity analogous to a Péclet number²³:

$$Pe = \frac{KL^{2m+1}}{D} \quad (2)$$

where L is a horizontal length scale. Setting $Pe = 1$, we solved for a characteristic length, L_c , at which the timescales for advection and diffusion are equal:

$$L_c = \left(\frac{D}{K} \right)^{\frac{1}{2m+1}} \quad (3)$$

By solving equation (1) numerically, we found that the valley spacing, λ , is proportional to L_c (Fig. 2).

Topographic analysis. We used topographic data to infer the value of L_c for the study sites. At equilibrium ($\partial z / \partial t = 0$) and on hilltops, where A and $|\nabla z|$ approach zero, equation (1) reduces to:

$$\frac{U}{D} \approx -\nabla^2 z_h \quad (4)$$

and thus U/D can be inferred from the Laplacian of elevation on hilltops, $\nabla^2 z_h$. Rearranging equation (1) with $\partial z / \partial t = 0$ and using equation (4) gives:

$$\frac{|\nabla z|}{\nabla^2 z - \nabla^2 z_h} = \frac{D}{K} A^{-m} \quad (5)$$

This implies a power-law relationship between drainage area A and the quantity $|\nabla z| / (\nabla^2 z - \nabla^2 z_h)$, which we abbreviate as S^* . We used laser altimetry data to calculate A , $|\nabla z|$, and $\nabla^2 z$, and measured $\nabla^2 z_h$ as the value that $\nabla^2 z$ approaches as $A|\nabla z| \rightarrow 0$ (Fig. 3a). We then calculated S^* , found D/K and m from least-squares regression of $\log_{10}(S^*)$ against $\log_{10}(A)$ (Fig. 3b), and calculated L_c using equation (3). Values of $\nabla^2 z_h$, D/K , m , L_c and λ for the five study sites are listed in Table 1.

Received 11 February; accepted 26 May 2009.

- Shaler, N. S. Spacing of rivers with reference to the hypothesis of base-leveelling. *Geol. Soc. Am. Bull.* **10**, 263–276 (1899).
- Izumi, N. & Parker, G. Inception of channelization and drainage basin formation: upstream-driven theory. *J. Fluid Mech.* **283**, 341–363 (1995).
- Hovius, N. Regular spacing of drainage outlets from linear mountain belts. *Basin Res.* **8**, 29–44 (1996).

- Talling, P. J., Stewart, M. D., Stark, C. P., Gupta, S. & Vincent, S. J. Regular spacing of drainage outlets from linear fault blocks. *Basin Res.* **9**, 275–302 (1997).
- Izumi, N. & Parker, G. Linear stability analysis of channel inception: downstream-driven theory. *J. Fluid Mech.* **419**, 239–262 (2000).
- Perron, J. T., Kirchner, J. W. & Dietrich, W. E. Spectral signatures of characteristic spatial scales and non-fractal structure in landscapes. *J. Geophys. Res.* **113**, F04003, doi:10.1029/2007JF000866 (2008).
- Willgoose, G., Bras, R. L. & Rodriguez-Iturbe, I. Results from a new model of river basin evolution. *Earth Surf. Process. Landf.* **16**, 237–254 (1991).
- Howard, A. D. A detachment-limited model of drainage basin evolution. *Wat. Resour. Res.* **30**, 2261–2286 (1994).
- Tucker, G. E. & Bras, R. L. Hillslope processes, drainage density, and landscape morphology. *Wat. Resour. Res.* **34**, 2751–2764 (1998).
- Gilbert, G. K. *Report on the Geology of the Henry Mountains* 99–150 (US Government Printing Office, 1877).
- Davis, W. M. The convex profile of badland divides. *Science* **20**, 245 (1892).
- Gilbert, G. K. The convexity of hilltops. *J. Geol.* **17**, 344–350 (1909).
- Smith, T. R. & Bretherton, F. P. Stability and the conservation of mass in drainage basin evolution. *Wat. Resour. Res.* **8**, 1506–1529 (1972).
- Loewenherz, D. S. Stability and the initiation of channelized surface drainage: a reassessment of the short wavelength limit. *J. Geophys. Res.* **96**, 8453–8464 (1991).
- Simpson, G. & Schlunegger, F. Topographic evolution and morphology of surfaces evolving in response to coupled fluvial and hillslope sediment transport. *J. Geophys. Res.* **108**, 2300, doi:10.1029/2002JB002162 (2003).
- Dunne, T. Formation and controls of channel networks. *Prog. Phys. Geogr.* **4**, 211–239 (1980).
- Kirby, M. J. in *Geomorphological Models: Theoretical and Empirical Aspects* (ed. F. Ahnert) 1–14 (Catena, 1987).
- Willgoose, G., Bras, R. L. & Rodriguez-Iturbe, I. A physical explanation of an observed link area-slope relationship. *Wat. Resour. Res.* **27**, 1697–1702 (1991).
- Tarboton, D. G., Bras, R. L. & Rodriguez-Iturbe, I. A physical basis for drainage density. *Geomorphology* **5**, 59–76 (1992).
- Howard, A. D. Badland morphology and evolution: Interpretation using a simulation model. *Earth Surf. Process. Landf.* **22**, 211–227 (1997).
- Moglen, G. E., Eltahir, E. A. B. & Bras, R. L. On the sensitivity of drainage density to climate change. *Wat. Resour. Res.* **34**, 855–862 (1998).
- Horton, R. E. Erosional development of streams and their drainage basins: hydrophysical approach to quantitative morphology. *Bull. Geol. Soc. Am.* **56**, 275–370 (1945).
- Perron, J. T., Dietrich, W. E. & Kirchner, J. W. Controls on the spacing of first-order valleys. *J. Geophys. Res.* **113**, F04016, doi:10.1029/2007JF000977 (2008).
- Hilley, G. E. & Arrowsmith, J. R. Geomorphic response to uplift along the Dragon's Back pressure ridge, Carrizo Plain, California. *Geology* **36**, 367–370 (2008).
- Fernandes, N. F. & Dietrich, W. E. Hillslope evolution by diffusive processes: the timescale for equilibrium adjustments. *Wat. Resour. Res.* **33**, 1307–1318 (1997).
- Prosser, I. P. & Dietrich, W. E. Field experiments on erosion by overland flow and their implication for a digital terrain model of channel initiation. *Wat. Resour. Res.* **31**, 2867–2876 (1995).
- Montgomery, D. R. & Dietrich, W. E. Where do channels begin? *Nature* **336**, 232–234 (1988).
- Montgomery, D. R. & Dietrich, W. E. Channel initiation and the problem of landscape scale. *Science* **255**, 826–830 (1992).
- Rinaldo, A., Dietrich, W. E., Rigon, R., Vogel, G. K. & Rodriguez-Iturbe, I. Geomorphological signatures of varying climate. *Nature* **374**, 632–635 (1995).
- Seidl, M. A. & Dietrich, W. E. in *Functional Geomorphology* (eds K. H. Schmidt & J. de Ploey) 101–124 (Catena, 1992).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by the National Science Foundation (J.T.P.), the Institute for Geophysics and Planetary Physics (J.W.K. and J.T.P.), and NASA (W.E.D. and J.T.P.). Laser altimetry for Gabilan Mesa was acquired and processed by NCALM (www.ncalm.org) with support from the National Center for Earth-Surface Dynamics (NCED). We thank the Orradre family of San Ardo, California, for granting access to their land. We thank the states of Pennsylvania and Utah for making laser altimetry data publicly available. We also thank K. Whipple for his review.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.T.P. (perron@mit.edu).

LETTERS

Light limitation of nutrient-poor lake ecosystemsJan Karlsson¹, Pär Byström², Jenny Ask², Per Ask², Lennart Persson² & Mats Jansson²

Productivity denotes the rate of biomass synthesis in ecosystems and is a fundamental characteristic that frames ecosystem function and management. Limitation of productivity by nutrient availability is an established paradigm for lake ecosystems^{1–3}. Here, we assess the relevance of this paradigm for a majority of the world's small, nutrient-poor lakes, with different concentrations of coloured organic matter^{4,5}. By comparing small unproductive lakes along a water colour gradient, we show that coloured terrestrial organic matter controls the key process for new biomass synthesis (the benthic primary production) through its effects on light attenuation. We also show that this translates into effects on production and biomass of higher trophic levels (benthic invertebrates and fish). These results are inconsistent with the idea that nutrient supply primarily controls lake productivity, and we propose that a large share of the world's unproductive lakes, within natural variations of organic carbon and nutrient input, are limited by light and not by nutrients. We anticipate that our result will have implications for understanding lake ecosystem function and responses to environmental change. Catchment export of coloured organic matter is sensitive to short-term natural variability and long-term, large-scale changes, driven by climate and different anthropogenic influences^{6,7}. Consequently, changes in terrestrial carbon cycling will have pronounced effects on most lake ecosystems by mediating changes in light climate and productivity of lakes.

Nutrient limitation of productivity has been a cornerstone for the understanding of lake ecosystems. This paradigm predicts that increased input of nutrients increases production at the basic trophic level (phytoplankton), with concomitant effects on higher trophic levels. Accordingly, comparative data from oligotrophic to hyper-eutrophic lakes show an increase not only in phytoplankton production⁸ but also in top consumer (fish) biomass and yield with increasing total phosphorus^{9,10} (Fig. 1a).

The nutrient limitation paradigm is mainly based on studies of the pelagic habitat, not accounting for the fact that a substantial part of the resources supporting top consumers can be generated in the benthic habitat^{11,12}. Benthic algae, limited primarily by light availability, dominate whole-lake primary production in clear-water and shallow lakes⁸. Increased nutrient supply in such lakes increases phytoplankton biomass at the expense of benthic primary production⁸. Light penetration is controlled by phytoplankton light absorption, so an increase in pelagic production may compensate for loss of benthic production, and nutrients thus become a good predictor of whole-lake productivity. However, most natural lakes worldwide have low nutrient concentrations (Fig. 1a and Supplementary Table 3) and low phytoplankton biomass, and the variation in light penetration to the benthic habitat is mainly controlled by variation in input of coloured terrestrial organic matter¹³. There are no studies on the extent to which light absorption by terrestrial organic matter causes loss of benthic algal production or how this affects total lake productivity. For example, increased nutrient input associated with

terrestrial organic matter¹⁴ can stimulate phytoplankton production and compensate for losses in benthic primary production, although it has been suggested that light absorption by coloured organic matter may even depress pelagic primary production^{13,15}. Thus, it is not known to what extent variation in productivity at different trophic levels depends on nutrient availability or light availability in lake ecosystems with low input of nutrients.

As a first test of the applicability of the nutrient limitation concept on small unproductive lakes, we compiled literature data on the relationship between fish biomass and total phosphorus in oligotrophic and mesotrophic (total phosphorus, 0–30 µg l⁻¹) lakes in Finland, New Zealand and Sweden. In contrast to the situation in

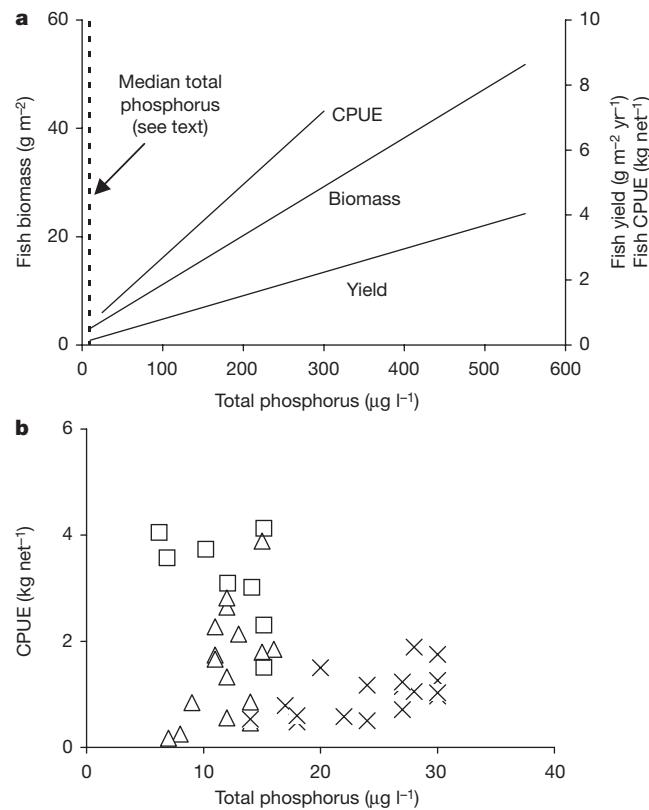


Figure 1 | Fish biomass and yield in temperate lakes. **a**, Published relationships ($r^2 = 0.75\text{--}0.84$) between fish biomass⁹, yield⁹ and catch per unit effort (CPUE)¹⁰. The vertical dashed line shows the average ($12 \mu\text{g l}^{-1}$) of reported median total phosphorus concentration in Norway ($2 \mu\text{g l}^{-1}$, $n = 1,006$), Finland ($13 \mu\text{g l}^{-1}$, $n = 873$), Sweden ($8 \mu\text{g l}^{-1}$, $n = 3,025$) and Wisconsin (United States) ($12 \mu\text{g l}^{-1}$, $n = 168$). **b**, Fish CPUE in oligotrophic and mesotrophic lakes (0–30 µg l⁻¹) from Finland (crosses), Sweden (triangles) and New Zealand (squares) as a function of total phosphorus. See Supplementary Tables 3 and 4.

¹Climate Impacts Research Centre (CIRC), Department of Ecology and Environmental Science, Umeå University, Box 62, SE-901 07 Abisko, Sweden. ²Department of Ecology and Environmental Science, Umeå University, SE-901 87 Umeå, Sweden.

more nutrient-rich lakes (Fig. 1a), we found no relationship between fish biomass and total phosphorus (Fig. 1b), suggesting that factors other than nutrients determine the development of fish biomass in these ecosystems. To elucidate the roles of nutrients and light as limiting factors for the productivity of naturally nutrient-poor lakes in detail, we then carried out a detailed study in 12 lake ecosystems where we measured production of algae and bacteria, the biomass of intermediate consumers, and the biomass and production of top consumers (fish). On a global scale, the selection of lakes represents the most common lake types^{4,5} in terms of small size (area, 0.02–0.17 km²; mean depth, 2.6–6.0 m), low nutrient concentrations (total phosphorus, 4.1–24.0 µg l⁻¹; total nitrogen, 89–483 µg l⁻¹) and range in dissolved organic carbon (DOC) concentrations (2.4–16.8 mg l⁻¹) (Supplementary Table 1). The DOC concentration, representing conditions for lakes in a clear- to brown-water gradient, was strongly related to lake water colour ($r^2 = 0.96, P < 0.001$) and to the vertical light attenuation of photosynthetically active radiation (PAR; $r^2 = 0.97, P < 0.001$). Terrestrial organic matter, which dominated the DOC pool in these lakes¹⁶, thus had a strong negative influence on light penetration, and mean light irradiance of the whole-lake volumes (I_m), expressed as a fraction of surface light, varied between 0.05 and 0.57.

We measured the basal production of algae and bacteria (primary production and heterotrophic bacterial production based on terrestrial organic carbon) in benthic and pelagic habitats¹⁷. These processes represent mobilization of energy in the lake from external sources and form a base for production at all trophic levels^{13,18,19}. The production (mean ± 1 s.d.) of benthic algae ($86 \pm 66 \text{ mg C m}^{-2} \text{ d}^{-1}$) showed high values and spanned a larger range between lakes compared to the production of phytoplankton ($11 \pm 8 \text{ mg C m}^{-2} \text{ d}^{-1}$) and heterotrophic bacteria based on terrestrial organic carbon in pelagic ($23 \pm 19 \text{ mg C m}^{-2} \text{ d}^{-1}$) and benthic ($14 \pm 14 \text{ mg C m}^{-2} \text{ d}^{-1}$) habitats in the lakes. The biomass of zoobenthos varied between 0.01 and 1.16 g m⁻² and was positively related ($r^2 = 0.49, P = 0.012$) to the benthic basal production. The biomass of zooplankton varied between 0.02 and 1.23 g m⁻² and was not related to pelagic basal production. All lakes had allopatric fish populations of either Arctic char (*Salvelinus alpinus*) or Eurasian perch (*Perca fluviatilis*), two of the most common top consumers in Arctic and boreal regions²⁰. From lake-specific, size-dependent growth functions and the size structure of the fish populations, we calculated the relative biomass production of fish to be between 30 and 565 g net⁻¹ yr⁻¹ in the lakes.

The whole-lake primary production ($r^2 = 0.81, P < 0.001$), basal production by algae and bacteria ($r^2 = 0.65, P = 0.001$) and production of top consumers ($r^2 = 0.59, P = 0.004$) was negatively related to total phosphorus in the lake water (Fig. 2). Therefore, factors other than nutrient supply controlled the biomass production in these lakes. The decrease in whole-lake basal production (from 234.8 down to 24.1 mg C m⁻² d⁻¹) with increasing nutrient content was related to the decrease in benthic primary production (from 195.8 down to 0.1 mg C m⁻² d⁻¹) along the gradient. Epipelagic algae, which dominate the basal production in clear lakes¹⁷, retrieve nutrients from the sediments, and their exploitation of these nutrients is mainly regulated by the light penetration to the sediment surface²¹. Accordingly, light attenuation and mean depth of the lakes explained 73% ($P < 0.001$) of the variability in whole-lake benthic primary production among the lakes. Increased production of phytoplankton and heterotrophic bacteria did not compensate for the loss of benthic primary production with decreasing light penetration, despite higher nutrient content in the more coloured lakes.

Analyses of stable carbon isotopes ($\delta^{13}\text{C}$) showed that the decrease in light and the consequent loss of benthic primary production also decreased the importance of this energy pathway for fish. The mean $\delta^{13}\text{C}$ of fish in the lakes ranged between -22.9‰ and -31.4‰ with higher values, similar to those of surface sediment and zoobenthos, in clear-water lakes and with lower values, more similar to those of

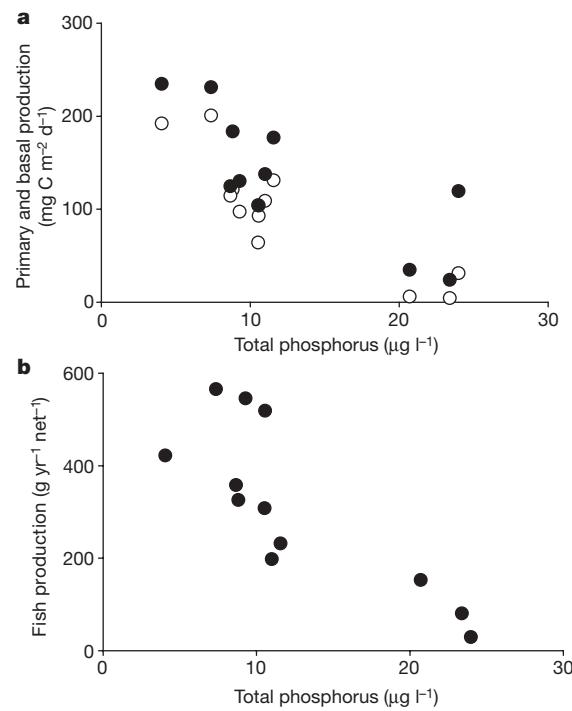


Figure 2 | Production as a function of nutrients. **a**, Whole-lake (benthic+pelagic) primary production (open circles) and basal production (filled circles, primary production plus bacterial production based on allochthonous organic carbon) as a function of total phosphorus. **b**, Whole-lake fish production as a function of total phosphorus. Solid line, error bars.

zooplankton, in humic lakes (Fig. 3a). The contribution of benthic algal carbon to fish biomass (F_{bent}) was estimated using a two-source mixing model, including the $\delta^{13}\text{C}$ of benthic (littoral zoobenthos) and pelagic (cladoceran zooplankton) energy sources, and showed that F_{bent} varied between 14% and 98% and was positively related ($r^2 = 0.86, P < 0.001$) to the mean light irradiance in the lakes (Fig. 3b).

The change in basal resource use by fish from benthic dominance in clear-water lakes to pelagic dominance in coloured humic lakes also affected the biomass production of fish. Fish production was positively related to benthic primary production ($r^2 = 0.49, P = 0.012$) and zoobenthos biomass ($r^2 = 0.45, P = 0.018$) in the lakes. Basal production by phytoplankton and heterotrophic bacteria did not correlate to fish production and did not significantly add to the explanation of fish production in the regression analysis (Supplementary Table 2). Furthermore, fish production was negatively related ($r^2 = 0.44, P = 0.019$) to zooplankton biomass in the lakes. Thus, an increase in fish resource use and growth on organic carbon generated in the pelagic habitat did not compensate for the loss of fish resource use and growth on autochthonous organic carbon generated in the benthic habitat. Consequently, we found that fish production was positively related ($r^2 = 0.63, P = 0.002$) to the annual light irradiance in the lakes (Fig. 4a), suggesting that light availability ultimately controlled fish production by its impact on benthic algal primary production.

We extended our analysis to include a larger data set (33 lakes in Sweden and Finland, Supplementary Table 4) to test the generality of our findings concerning the relationship between light irradiance and fish production. This analysis confirmed the fundamental light–fish relationship using annual light irradiance and fish biomass (Fig. 4b, all lakes: $r^2 = 0.52, P < 0.001$). Our results stress the crucial role of light availability in controlling not only basal but also top consumer production and biomass. Furthermore, our results show how interactions across ecosystem boundaries (that is, terrestrial organic matter controls lake productivity) and habitat boundaries (that is, exploitation of benthic and pelagic resources by top consumers) determine the

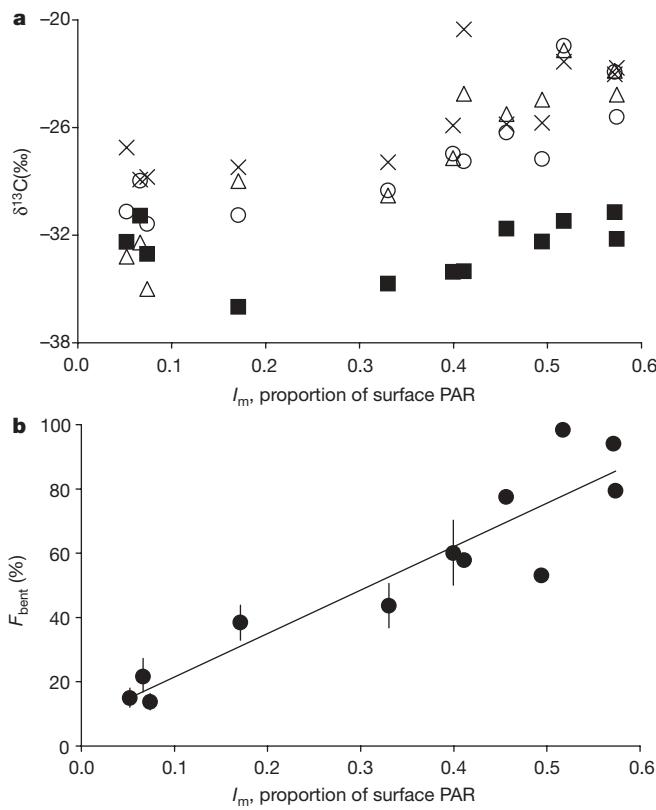


Figure 3 | Fish resource use. **a**, The mean $\delta^{13}\text{C}$ of fish (circles), shallow surface sediment (crosses), cladocerans zooplankton (squares) and chironomid zoobenthos (triangles) as a function of the mean light irradiance (I_m) in the lakes. **b**, Benthic contributions to fish body carbon (F_{bent}) as a function of I_m . The solid line describes the linear relationship between F_{bent} and I_m ($r^2 = 0.86, P < 0.001$). Error bars (s.d.) in the six lakes where littoral zoobenthos samples were not available for isotope analysis represent the uncertainty introduced by calculating F_{bent} using the range in $\delta^{13}\text{C}$ of littoral zoobenthos from other lakes.

production of lake ecosystems. The negative effect on lake productivity by terrestrial carbon also indicates that this carbon represents a repression of lake productivity rather than a subsidy.

We question the general value of the nutrient limitation paradigm for explaining and predicting the productivity of unproductive lake ecosystems. Undoubtedly, input of nutrients resulting, for example, from cultural eutrophication can stimulate pelagic biomass production of nutrient-poor lakes³. However, we show that changes in nutrient input do not necessarily result in increased total ecosystem productivity. Nutrients input in unproductive lakes is normally associated with terrestrial organic matter¹⁴, and the effects of terrestrial organic matter on light penetration and benthic (light-limited) photosynthesis override possible positive effects of nutrients on pelagic (nutrient-limited) production. We suggest that nutrient availability is a major controlling factor of lake productivity mainly in systems that are dominated by pelagic production (for example, eutrophic lakes or very deep oligotrophic lakes), whereas light availability determines the productivity in a majority of the world's unproductive lakes within natural variations of organic matter and nutrient input. The light climate of most unproductive lakes is a function of lake depth and, more importantly, of coloured organic matter input from the catchment. Catchment export of organic matter and related lake DOC concentrations in northern (temperate, boreal and subarctic) areas where most of the world's unproductive lakes are located varies within one to two orders of magnitude²². This variation is largely related to differences in catchment vegetation, air temperature and runoff^{6,23}. Consequently, spatial and temporal variation in lake productivity determined by light absorption of organic material, as

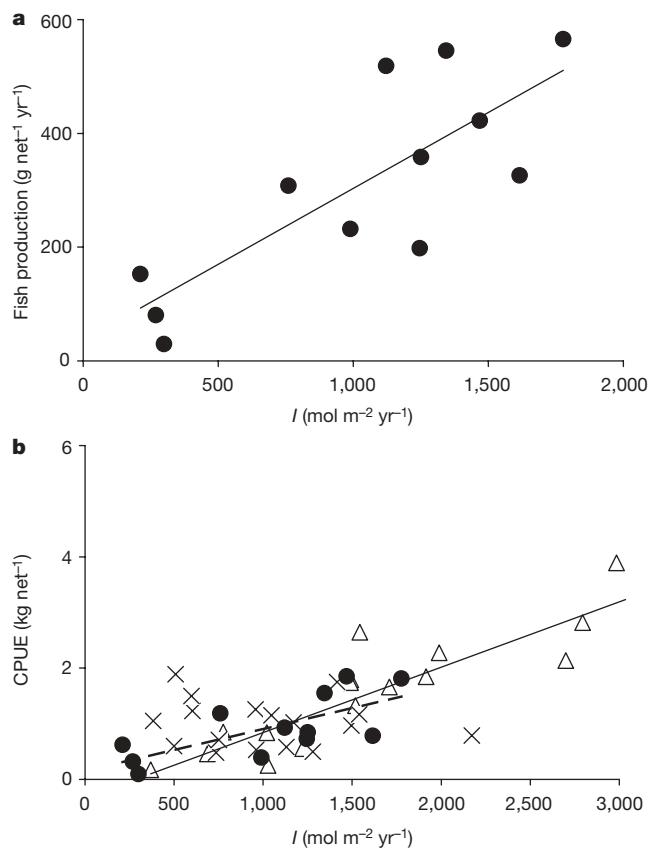


Figure 4 | Fish production and biomass as a function of light. **a**, Fish production as a function of the annual light climate (I , representing the mean PAR in the whole-lake volume during the ice-free period) in the 12 lakes ($r^2 = 0.63, P = 0.002$). **b**, Fish CPUE as a function of I in the 12 study lakes (circles, $r^2 = 0.50, P = 0.010$, dashed line) and in 33 additional lakes ($r^2 = 0.50, P < 0.001$, solid line) from Finland (crosses) and Sweden (triangles). There was no difference ($P = 0.76$) in the slope between the two regression lines. For references see Supplementary Table 4.

shown in this study, should be pronounced (Fig. 4) and also susceptible to changes in catchment carbon cycling induced by, for example, climate and land use change.

METHODS SUMMARY

We sampled 12 lakes in northern Sweden ($64\text{--}68^\circ \text{N}, 18\text{--}19^\circ \text{E}$) in midsummer. A composite water sample was collected and analysed for DOC, total nitrogen, total phosphorus and absorbance (440 nm)¹⁷. The vertical attenuation coefficient (K_d) was calculated from the slope of the linear regression of the natural logarithm of PAR versus depth. The mean light irradiance (I_m) in the lakes was calculated as the integrated light of the whole-lake volume, and expressed as a fraction of surface light (PAR at depth 0). Annual light irradiance in the lakes was calculated by multiplying I_m with surface PAR during the ice-free period.

Pelagic primary production was measured (five to eight depths, one occasion) using the ^{14}C method. Benthic primary production was obtained (one occasion) from changes in dissolved inorganic carbon in water overlying sediment cores (six depths) or stones (one depth)¹⁷. Bacterial production was measured on composite water (pelagic) and sediment (benthic) samples, using the $[^3\text{H}]$ leucine method^{17,24}. Fish were sampled with multi-mesh gillnets in benthic and pelagic habitats (two to four per habitat). From length-age and weight-length relationships we derived size-dependent yearly growth rate functions for each lake. The fish production per net was calculated as the sum of individual yearly growth of all captured fish divided by number of gillnets used. Zooplankton was sampled (two to four occasions, three locations) using a plankton net, zoobenthos was sampled (one occasion, three locations) using an Ekman dredge, and the biomass was calculated using length-weight regressions^{25,26}. All data were converted to integrated whole-lake values using bathymetric data of the lakes¹⁷. We measured $\delta^{13}\text{C}$ of chironomids (two to five samples per lake), cladocerans (two to five samples per lake), surface sediment (five to eight samples per lake) and fish dorsal muscle (25–30 samples per lake) at the Department of Geology and Geochemistry, Stockholm University.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 25 March; accepted 28 May 2009.

1. Schindler, D. W. Evolution of phosphorus limitation in lakes. *Science* **195**, 260–262 (1977).
2. Sterner, R. W. On the phosphorus limitation paradigm for lakes. *Int. Rev. Hydrobiol.* **93**, 433–445 (2008).
3. Carpenter, S. R. Phosphorus control is critical to mitigating eutrophication. *Proc. Natl Acad. Sci. USA* **105**, 11039–11040 (2008).
4. Downing, J. A. et al. The global abundance and size distribution of lakes, ponds, and impoundments. *Limnol. Oceanogr.* **51**, 2388–2397 (2006).
5. Sobek, S., Tranvik, L. J., Prairie, Y. T., Kortelainen, P. & Cole, J. J. Patterns and regulation of dissolved organic carbon: An analysis of 7,500 widely distributed lakes. *Limnol. Oceanogr.* **52**, 1208–1219 (2007).
6. Tranvik, L. J. & Jansson, M. Climate change — terrestrial export of organic carbon. *Nature* **415**, 861–862 (2002).
7. Roulet, N. & Moore, T. R. Environmental chemistry: browning the waters. *Nature* **444**, 283–284 (2006).
8. Vadéboncoeur, Y. et al. From Greenland to green lakes: cultural eutrophication and the loss of benthic pathways in lakes. *Limnol. Oceanogr.* **48**, 1408–1418 (2003).
9. Hanson, J. M. & Leggett, W. C. Empirical predictions of fish biomass and yield. *Can. J. Fish. Aquat. Sci.* **39**, 257–263 (1982).
10. Jeppesen, E. et al. Lake and catchment management in Denmark. *Hydrobiologia* **395/396**, 419–432 (1999).
11. Vander Zanden, M. J. & Vadéboncoeur, Y. Fishes as integrators of benthic and pelagic food webs in lakes. *Ecology* **83**, 2152–2161 (2002).
12. Karlsson, J. & Byström, P. Littoral energy mobilization dominates energy supply for top consumers in subarctic lakes. *Limnol. Oceanogr.* **50**, 538–543 (2005).
13. Jones, R. I. The influence of humic substances on lacustrine planktonic food-chains. *Hydrobiologia* **229**, 73–91 (1992).
14. Meili, M. Sources, concentrations and characteristics of organic matter in softwater lakes and streams of the Swedish forest region. *Hydrobiologia* **229**, 23–41 (1992).
15. Carpenter, S. R., Cole, J. J., Kitchell, J. F. & Pace, M. L. Impact of dissolved organic carbon, phosphorus, and grazing on phytoplankton biomass and production in experimental lakes. *Limnol. Oceanogr.* **43**, 73–80 (1998).
16. Karlsson, J., Jonsson, A., Meili, M. & Jansson, M. Control of zooplankton dependence on allochthonous organic carbon in humic and clear-water lakes in northern Sweden. *Limnol. Oceanogr.* **48**, 269–276 (2003).
17. Ask, J. et al. Whole-lake estimates of carbon flux through algae and bacteria in benthic and pelagic habitats of clear-water lakes. *Ecology* **90**, 1923–1932 (2009).
18. Jansson, M., Persson, L., De Roos, A. M., Jones, R. I. & Tranvik, L. J. Terrestrial carbon and intraspecific size-variation shape lake ecosystems. *Trends Ecol. Evol.* **22**, 316–322 (2007).
19. Carpenter, S. R. et al. Ecosystem subsidies: Terrestrial support of aquatic food webs from ^{13}C addition to contrasting lakes. *Ecology* **86**, 2737–2750 (2005).
20. Svärdson, G. *Interspecific Population Dominance in Fish Communities of Scandinavian Lakes*. Report No. 55, 144–171 (Institute of Freshwater Research, 1976).
21. Hansson, L.-A. Factors regulating periphytic algal biomass. *Limnol. Oceanogr.* **37**, 322–328 (1992).
22. Mulholland, P. J. in *Aquatic Ecosystems: Interactivity of Dissolved Organic Matter* (eds Findlay, S. & Sinsabaugh R. L.) Ch. 6, 139–159 (Academic, 2003).
23. Hope, D., Billett, M. F. & Cresser, M. S. A review of the export of carbon in river water: fluxes and processes. *Environ. Pollut.* **84**, 301–324 (1994).
24. Karlsson, J., Jansson, M. & Jonsson, A. Similar relationships between pelagic primary and bacterial production in clearwater and humic lakes. *Ecology* **83**, 2902–2910 (2002).
25. McCauley, E. in *A Manual on Methods for the Assessment of Secondary Productivity in Fresh Waters* 2nd edn (eds Downing, J. A. & Riegler, F. H.) 228–265 (Blackwell Scientific, 1984).
26. Persson, L., Andersson, J., Wahlstrom, E. & Eklov, P. Size-specific interactions in lake systems: predator gape limitation and prey growth rate and mortality. *Ecology* **77**, 900–911 (1996).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The study was part of the research environment LEREC (Lake Ecosystem Response to Environmental Change), financially supported by Formas (the Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning). We thank the Abisko Scientific Research Station, the Swedish Meteorological and Hydrological Institute and the Umeå Marine Sciences Centre for providing PAR data, M. Mörtö and J. Johansson for chemical analysis and C. Hein for language editing.

Author Contributions J.K., P.B., L.P. and M.J. contributed to study design. J.K., P.B., J.A. and P.A. contributed to sampling and analysis of data. J.K. and M.J. wrote the paper. All authors commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.K. (Jan.Karlsson@emg.umu.se).

METHODS

Calculations. Water colour was calculated from absorbance at 440 nm (ref. 27). The mean light irradiance of the whole-lake volume (I_m), that is, a dimensionless estimate that expresses the variation in light climate between lakes, was calculated as:

$$I_m = (1 - e^{K_d \times Z_m}) / (K_d \times Z_m)$$

where K_d is the vertical attenuation coefficient and Z_m the mean depth of the lakes. The annual light irradiance (I) in the lakes was calculated as the product of I_m and the surface PAR during the ice-free period.

The contribution of benthic algae to fish (F_{bent}) was estimated using a two-source mixing model including the stable carbon isotopic signature of benthic ($\delta^{13}\text{C}_{bent}$) and pelagic ($\delta^{13}\text{C}_{pel}$) energy sources and a trophic fractionation factor (F) of 0.4‰ as:

$$F_{bent} = (\delta^{13}\text{C}_{\text{fish}} - F - \delta^{13}\text{C}_{\text{pel}}) / (\delta^{13}\text{C}_{\text{bent}} - \delta^{13}\text{C}_{\text{pel}})$$

Cladoceran zooplankton was used as the pelagic end-member¹⁶, and littoral zoobenthos (chironomids) was used as the benthic end-member¹². The model assumes that cladocerans and chironomid zooplankton are consumed directly by fish.

The littoral sediments were defined as the benthic habitat where primary production exceeded respiration (that is, net autotrophic with negligible respiration of external organic carbon). Littoral zoobenthos were obtained from six lakes and in the other lakes F_{bent} was calculated using the mean $\delta^{13}\text{C}$ of littoral zoobenthos from the other lakes (Supplementary Fig. 1). This introduced an uncertainty in calculated F_{bent} . The effect of variation in the $\delta^{13}\text{C}$ of the benthic end-member in these six lakes was tested by calculating F_{bent} using the range of $\delta^{13}\text{C}$ of the benthic end-member. The $\delta^{13}\text{C}$ are calculated as:

$$\delta^{13}\text{C} = (R_{\text{sample}}/R_{\text{standard}} - 1) \times 1,000$$

where $R = {}^{13}\text{C}/{}^{12}\text{C}$.

We calculated total new basal biomass production in both pelagic and benthic habitats as the sum of measured autotrophic production and calculated

heterotrophic bacterial production based on allochthonous organic carbon¹⁷. Bacterial production based on autochthonous organic carbon represents recycling of material and was not included in the measure of new basal biomass production. Total basal biomass production for the whole lake was obtained as the sum of benthic and pelagic production.

Literature data. Published relationships between total phosphorus and fish CPUE, biomass and yield were obtained from the literature. A summary of the literature data is given in Supplementary Table 3.

We also collected literature data on fish CPUE, total phosphorus and I . A selection of lakes was based on total phosphorus (0–30 µg l⁻¹), standardized sampling of fish (multi-mesh nets) and data for calculating I . A summary of the literature data are given in Supplementary Table 4. CPUE was standardized based on net area and reported as kilograms per net (30 × 1.5 m) per night. The annual light irradiance in the lakes was calculated as above when possible. When not directly measured, K_d was estimated from data on lake water colour or Secchi depth using relationships between K_d and colour (this study) and between colour and Secchi depth²⁸. Surface PAR data for the published lakes were obtained from the STRÅNG model for solar radiation, derived from the Swedish Meteorological and Hydrological Institute, and were produced with support from the Swedish Radiation Protection Authority and the Swedish Environmental Agency. The length of the ice-free season, when not stated, was obtained from the literature²⁹.

Statistical analysis. Variables were log transformed to obtain normality. The data ($n = 12$) were analysed using least-squares linear regression, partial correlation (Pearson) and Akaike's information criterion corrected for small sample sizes³⁰. Slopes of linear regressions on the lakes from this study ($n = 12$) and from the literature ($n = 33$) in Fig. 4b were compared using a two-tailed *t*-test.

27. Cuthbert, I. D. & del Giorgio, P. Towards a standard method of measuring color in freshwater. *Limnol. Oceanogr.* **37**, 1319–1326 (1992).
28. Eloranta, P. Light penetration in different types of lakes in Central Finland. *Holarct. Ecol.* **1**, 362–366 (1978).
29. Blenckner, T., Jarvinen, M. & Weyhenmeyer, G. A. Atmospheric circulation and its impact on ice phenology in Scandinavia. *Boreal Environ. Res.* **9**, 371–380 (2004).
30. McQuarrie, A. D. R. & Tsai, C.-L. *Regression and Time Series Model Selection* (World Scientific, 1998).

LETTERS

Partial penetrance facilitates developmental evolution in bacteria

Avigdor Eldar^{1*}, Vasant K. Chary^{2*}, Panagiotis Xenopoulos², Michelle E. Fontes¹, Oliver C. Losón¹, Jonathan Dworkin³, Patrick J. Piggot² & Michael B. Elowitz¹

Development normally occurs similarly in all individuals within an isogenic population, but mutations often affect the fates of individual organisms differently^{1–4}. This phenomenon, known as partial penetrance, has been observed in diverse developmental systems. However, it remains unclear how the underlying genetic network specifies the set of possible alternative fates and how the relative frequencies of these fates evolve^{5–8}. Here we identify a stochastic cell fate determination process that operates in *Bacillus subtilis* sporulation mutants and show how it allows genetic control of the penetrance of multiple fates. Mutations in an intercompartmental signalling process generate a set of discrete alternative fates not observed in wild-type cells, including rare formation of two viable ‘twin’ spores, rather than one within a single cell. By genetically modulating chromosome replication and septation, we can systematically tune the penetrance of each mutant fate. Furthermore, signalling and replication perturbations synergize to significantly increase the penetrance of twin sporulation. These results suggest a potential pathway for developmental evolution between monosporulation and twin sporulation through states of intermediate twin penetrance. Furthermore, time-lapse microscopy of twin sporulation in wild-type *Clostridium oceanicum* shows a strong resemblance to twin sporulation in these *B. subtilis* mutants^{9,10}. Together the results suggest that noise can facilitate developmental evolution by

enabling the initial expression of discrete morphological traits at low penetrance, and allowing their stabilization by gradual adjustment of genetic parameters.

Under nutrient-limited conditions, an individual *B. subtilis* cell can develop into a resilient dormant spore¹¹. Many sporulation mutations reduce the fraction of cells that sporulate successfully (Fig. 1a, b)^{4,12–14}. This makes sporulation an ideal model system in which to study the origins and impact of partial penetrance.

At the onset of sporulation, *B. subtilis* cells divide asymmetrically into smaller (forespore) and larger (mother-cell) compartments. Septation leads to the forespore-specific activation of the transcriptional regulator σ^F (Fig. 1c, d)¹¹. This in turn activates expression of *spoIIR*, which initiates an intercompartmental signalling cascade that activates the mother-cell-specific regulator σ^E , causing mother-cell differentiation^{11,15}. Deleting σ^E allows a second asymmetric septum to form, resulting in ‘abortively disporic’ cells with two DNA-containing immature forespores and a mother cell devoid of DNA^{16,17}. Attenuation of *spoIIR* expression results in a partially penetrant mixture of successfully sporulating and abortively disporic cells^{12,13}.

To explore the effects of *spoIIR* mutations on sporulation penetrance, we constructed a set of strains, collectively denoted as *spoIIR^{PP}* mutants, in which the rate and/or the time of onset of *spoIIR* expression is specifically perturbed (Supplementary Methods

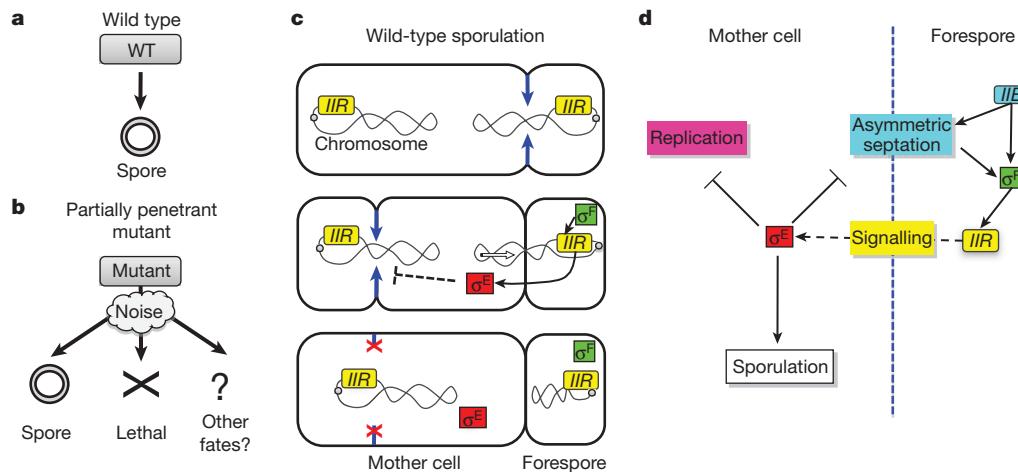


Figure 1 | Partial penetrance in the developmental process of sporulation.

a, In wild-type sporulation, each sporulating cell produces a single spore.
b, Partly penetrant mutants exhibit a mixture of normal sporulation, lethal

failures (cross) and alternative viable fates (?) due to cellular fluctuations (cloud). **c, d**, Events (**c**) and genetic interactions (**d**) leading to differentiation of the mother-cell and forespore compartments (see text).

¹Howard Hughes Medical Institute and Division of Biology and Department of Applied Physics, California Institute of Technology, Pasadena, California 91125, USA. ²Department of Microbiology and Immunology, Temple University School of Medicine, 3400 North Broad Street, Philadelphia, Pennsylvania 19140, USA. ³Department of Microbiology, College of Physicians and Surgeons, Columbia University, New York, New York 10032, USA.

*These authors contributed equally to this work.

and Supplementary Figs 1–4). We characterized *spoIIR^{PP}* mutants using time-lapse imaging of cells expressing fluorescent reporters of σ^F and σ^E activity (Methods, Supplementary Fig. 1 and Supplementary Movie 1). These movies revealed a diverse set of discrete cell fates (Fig. 2), whose relative frequencies depended on the type and severity of the perturbation to *spoIIR* expression. In all of these mutants, after an initial asymmetric septation and activation of σ^F in the forespore, mother cells exhibited one of three ‘primary’ fates (Fig. 2a–c). One population of cells activated σ^E and continued to sporulate normally (Fig. 2a). Another population formed the abortively disporic morphology (Fig. 2b). In the third population, neither σ^E activation nor asymmetric septation was observed (Supplementary Fig. 5). In these cells, the activated forespores did not develop further, but the mother cells continued to grow in a process we term ‘sporulation escape’ (Fig. 2c, Supplementary Fig. 5 and Methods)^{11,18}.

Escape enabled the formation of an additional cell type. About 25% of the time, escaping cells immediately re-initiated sporulation. In ~5% of such cases, two new forespores developed sequentially (~20 min apart) within the same mother cell (Fig. 2d and Supplementary Fig. 6). Unlike abortively disporic cells, these ‘twin’ sporulating cells completed sporulation, producing two mature viable spores in a single mother-cell compartment (Supplementary Fig. 7). Twins exhibited the transcriptional and morphological hallmarks of proper sporulation (Fig. 2d and Supplementary Fig. 8). Twin spores germinated properly, and were ultraviolet-resistant (Supplementary Fig. 9). Similar morphologies (called ‘bipolar’) appear to have evolved independently many times in the class Clostridia^{9,10}, the sporulation pathway of which is homologous to that of *Bacillus*¹⁹, but have not been observed in *Bacillus* itself (Supplementary Fig. 10). Twin sporulation may be adaptive under some conditions, including when vegetative growth is inhibited and proliferation occurs principally by sporulation^{10,20}.

The ability of cells to form twins is surprising because only two chromosome copies are present during normal sporulation¹¹, and twins require at least three. Therefore, we tracked the number and cellular compartment of chromosomes tagged with TetR-green fluorescent protein (GFP) ‘dots’ bound to a cassette of chromosomally integrated *tetO* operators (Fig. 2e and Supplementary Fig. 11)^{21,22}. We found that 30% of ‘escaping’ *spoIIR* mutant cells over-replicated to produce three or more chromosomal dots within a single cell. Fifteen per cent of these cells then underwent two consecutive asymmetric septation events without additional replication, producing twins. Because *spoIIR^{PP}* mutations cannot affect chromosome number before the initial septation event (when *spoIIR* is first expressed), this result explains why twins only formed as secondary fates in *spoIIR^{PP}* mutants. Furthermore, some of these over-replicating (polyploid) cells sporulated normally (to produce a single spore), despite the presence of extra chromosomal copies in the mother cell (Fig. 2f and Supplementary Fig. 11).

What determines the fate of an individual cell within a clonal mutant population? Because *spoIIR* mutations have alternative fates, and because sporulation penetrance is directly linked to the strength of *spoIIR* mutations^{12,13} (Supplementary Figs 3 and 4), fluctuations in *spoIIR* expression represent the most direct candidate for fate determination. To test this hypothesis, we constructed a specific *spoIIR^{PP}* strain, called *spoIIR^{PP-CY}* (Supplementary Fig. 2), in which *spoIIR* expression is reduced and delayed by ~10 min, and can be monitored using a co-transcribed *yfp* reporter. The onset of *spoIIR* expression can be compared in the same cell with that of another, non-delayed, σ^F -dependent promoter controlling *cfp* expression (Fig. 3a and Supplementary Fig. 2). We observed variation of 5 min in the timing and ~56% in the rate of *spoIIR* expression ($n = 148$ cells; Fig. 3b, inset). However, *spoIIR* expression rate fluctuations explained only ~15% of the decision between sporulation and other fates (based on relative

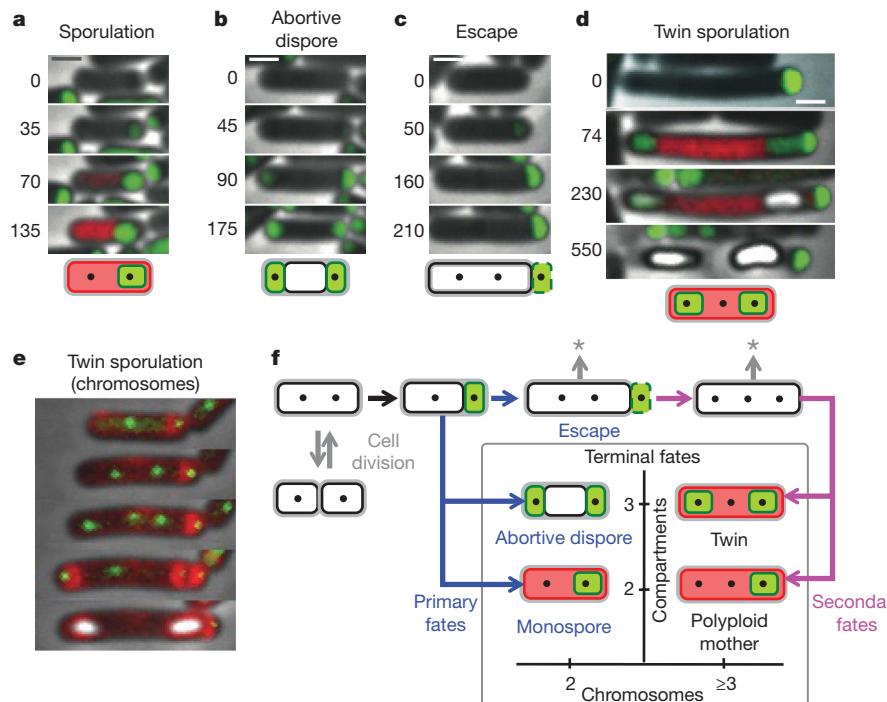


Figure 2 | Time-lapse movies reveal alternative developmental pathways in *spoIIR^{PP}* signalling mutants. **a–d**, Green and red represent fluorescent-protein expression from σ^F - and σ^E -dependent promoters, respectively, overlaid on phase-contrast images (grey). Developing forespores appear white at late times. Times are indicated in minutes from σ^F activation. **a**, Normal sporulation. **b**, Abortively disporic cells. **c**, Escaping cells activate σ^F but continue to elongate without activating σ^E (Supplementary Fig. 5). Note that the activated forespore (right) does not develop further. **d**, Twin sporulation occurs after escape. Green fluorescence at the initial time is a

remnant of escape from the previous sporulation attempt. **e**, Chromosome over-replication occurs before the formation of twin forespores. TetR-GFP-tagged chromosomal loci appear as green dots. Membrane staining (red) shows septation events. The rightmost dot is the remnant from a previous escape. **f**, Temporal sequence of events leading to observed terminal fates, which are classified by the numbers of chromosomes (x axis) and compartments (y axis). An asterisk indicates potential for return to vegetative division and/or additional sporulation attempt. Scale bar, 1 μ m.

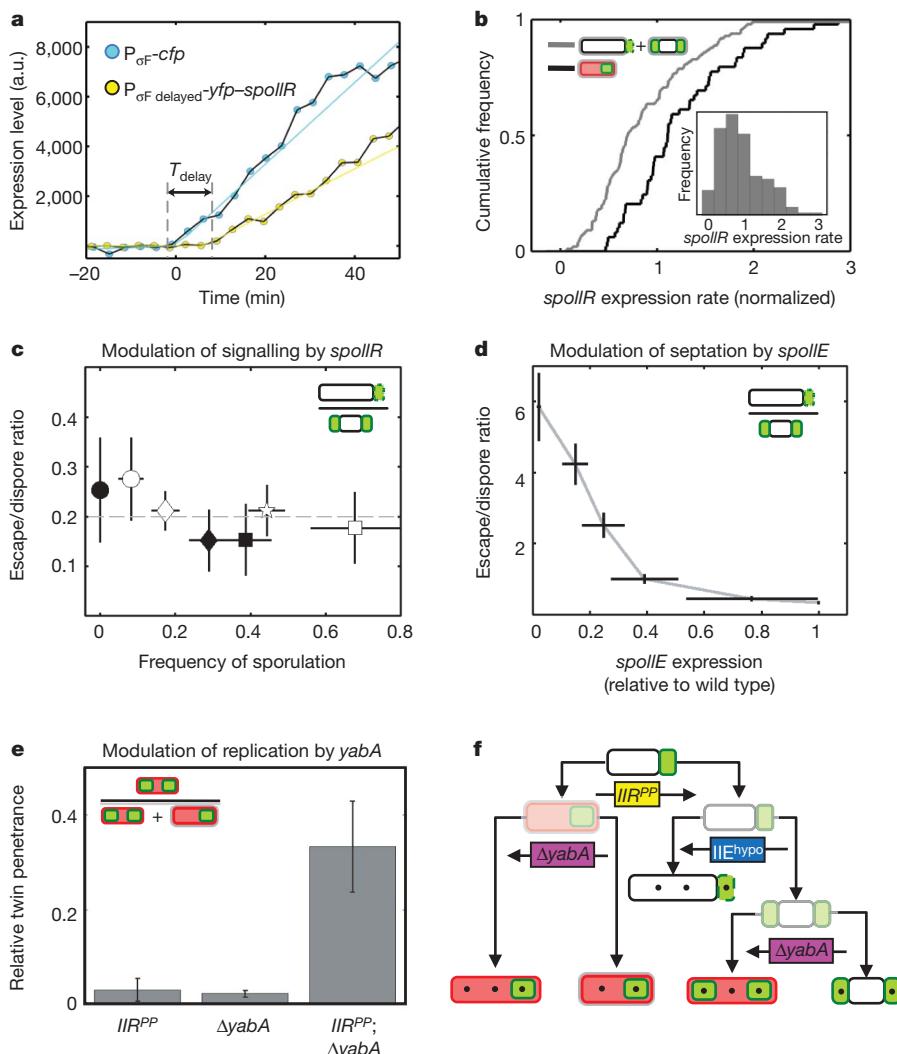


Figure 3 | Noise and gene expression control cell fate in a hierarchical fashion. **a**, Time traces indicating delay (T_{delay} , arrow) and reduction (slope of yellow line compared with that of cyan line) in *spoIIR* expression rate of a typical *spoIIR^{PF-CY}* cell. P_{σ_F} , σ^F -responsive promoter (Supplementary Methods); a.u., arbitrary units. **b**, Cumulative histograms of *spoIIR* expression rate are shown for two subpopulations of a single *spoIIR^{PP}* strain in the same microcolony ($n = 150$ cells). Sporulating cells show a systematically higher level of *spoIIR* expression. Inset, cell-cell variability in *spoIIR* expression rate. **c–e**, Systematic genetic manipulation of fate penetrance. Error bars (s.e.) are based on three replicate experiments. **c**, *spoIIR* expression controls the overall frequency of sporulation (x axis) but does not systematically affect the ratio of escape cells and abortive disporic cells (y axis). Points represent *spoIIR^{PP}* strains differing in *spoIIR* expression level and delay. Symbols correspond to different *spoIIR^{PP}* mutant strains, as defined in Supplementary Methods. Dashed line indicates the mean escape/dispore ratio across the measured strains. **d**, *spoIIIE* expression level tunes the penetrance ratio of escape and abortively disporic fates (Methods and Supplementary Fig. 12). **e**, Deletion of *yabA* interacts synergistically with *spoIIR^{PP}* mutants to increase twin penetrance (see also Supplementary Fig. 11). **f**, Fate determination can be controlled hierarchically: different genes affect different decision points.

mutual information; $P < 10^{-4}$). Variation in the expression delay had weaker explanatory power. Thus, fluctuations directly related to the genetic perturbation in *spoIIR* only partially account for cell fate. Much of the fate decision is apparently determined by other fluctuations, whose effects are revealed when *spoIIR* expression is attenuated.

Next we analysed the genetic control of fate penetrance in the *spoIIR*^{PP} mutants. Sporulation frequency varied from 0% in a Δ *spoIIR* mutant to approximately 75% across the set of *spoIIR*^{PP} strains. Strikingly, however, the ratio of escape and abortively disporig frequencies remained approximately constant across this range (Fig. 3c and Supplementary Methods). This behaviour suggests that two independent levels of primary cell fate determination can be distinguished: the decision between sporulating and non-sporulating fates depends on how *spoIIR* is perturbed, whereas the decision between escape and abortively disporig fates is independent of this perturbation (Fig. 3f).

Escape and abortively disporic fates differ in the number of asymmetric divisions they undergo and their ratio therefore may be affected by regulators of asymmetric division. The membrane protein SpoIIE promotes asymmetric septum formation²³, and therefore could have such an effect (Fig. 1d). To test this hypothesis, we constructed a strain, *spoIIE*^{hyp0}, in which the level of *spoIIE* expression could be modulated by the inducer isopropyl β-D-1-thiogalactopyranoside (IPTG) without affecting its temporal dynamics (Methods and Supplementary Fig. 12). This strain included a *ΔspoIIR* mutation to make the measurement independent of the additional indirect function of SpoIIE in *spoIIR* activation¹¹ (Fig. 1d). As expected, the ratio of

escape and abortively disporic frequencies increased as the level of *spoIIIE* expression was reduced (Fig. 3d and Supplementary Fig. 12).

Similarly, to increase the frequency of twin sporulation would require increasing both the number of chromosomes and the number of compartments generated during sporulation (Fig. 2f). Perturbations of *spoIIR* expression primarily affect compartment number, but have only a modest effect on chromosome number (through the escape state). We reasoned that mutations that increase chromosome copy number could synergize with *spoIIR*^{PP} mutations, increasing the frequency of twins at the expense of abortively disporic cells. To test this possibility, we used two mutations that increase chromosome replication: a null mutation of the chromosome replication inhibitor YabA²⁴ and a hypomorphic fusion of chromosome replication regulator Spo0J with GFP²² (Supplementary Fig. 11). Intriguingly, these over-replication mutations generated a low penetrance of twins in the wild-type background (Fig. 3e), suggesting that they may affect intercompartmental signalling. Combining these mutations with *spoIIR*^{PP} mutants permitted twins, as well as polyploid mother cells, to occur as a primary fate, because some cells then possessed additional chromosomes before the first asymmetric septation event (Supplementary Fig. 11). Finally, the over-replication mutations synergized with signalling mutations, increasing the penetrance of the twin fate to >30% of all viable sporangia, a level comparable to those observed in some natural twin-forming species²⁵ (Fig. 3e and Supplementary Fig. 11). Together, these results indicate, first, that all four terminal fates can be ‘primary’ (that is, occur without a round of escape) and, second, that the penetrance of all fates can

be manipulated by specific genetic perturbations affecting the processes of signalling, septation, and replication (Fig. 3f).

At the evolutionary level, partial penetrance could facilitate transitions between discrete phenotypes by enabling gradual changes in their frequencies, rather than an all-or-none switch from one phenotype to the other, which might require simultaneous changes in multiple processes. As described above, mutations that increase the probability of either additional septation or replication increase the penetrance of the maladaptive abortive-dispore state or the polyploid monospore state, respectively (Figs 2f and 3).

Two conditions would thus facilitate the gradual evolution of twin sporulation. First, a relatively high sporulation efficiency for the polyploid monospore state would allow sequential accumulation of mutations that favour over-replication to be followed by additional mutations that affect septation. We found that polyploid cells carry a fitness cost far smaller than that carried by abortive dispores, completing sporulation successfully about 75% as often as normal sporulating cells (Supplementary Fig. 11). Thus, there does not appear to be any fundamental obstacle to high fitness for this state. Second, mutations that cause correlation between replication and septation would favour the monospore and twin fates over the polyploid and abortive-dispore states. Analysis of fate frequencies revealed evidence for these correlations and their genetic control: in *spoIIR^{PP}* *spo0J-gfp* mutants, the frequency of reseptation (forming three compartments) differs by a factor of ~3 depending on the number of chromosomes observed before asymmetric septation ($P < 10^{-3}$; Fig. 4a). This correlation enhances the penetrance of twins and reduces the penetrance of the maladaptive abortive-dispore state. Neither over-replication nor reseptation is known to occur in wild-type strains; as a result, this correlation may not be under selection, and could therefore vary between backgrounds. Consistent with this, the measured interaction between *spoIIR^{PP}* and *yabA* mutations differed significantly between two closely related strain backgrounds (Fig. 4b). Together, these results suggest that under appropriate selection for twins, *B. subtilis* could acquire mutations that affect septation, replication and their correlation, and thereby enable gradual evolution of twin sporulation.

Finally, we sought to determine whether natural twin sporulation occurs in a manner similar to that observed in *B. subtilis* mutants. We acquired time-lapse movies of *C. oceanicum*, a marine anaerobe that exhibits twin sporulation, with fluorescent-membrane and DNA probes (Supplementary Movie 2)⁹. We observed a partially penetrant mixture of fates including twins and monospores, but not the abortively disporic fate. Analysis of DNA content in monospores was consistent with a subpopulation of polyploid mother cells, as in *B. subtilis* (Supplementary Fig. 13). We analysed the temporal sequence of septation and chromosome replication events in individual cells (Fig. 4c, Supplementary Movie 2 and Methods). DNA staining was observed to increase before the first septation event and

subsequently decrease in the mother cell (Fig. 4e), consistent with replication before the first septation and subsequent translocation of DNA into the forespores, as occurs in *B. subtilis* mutants. The two forespore compartments were formed by consecutive septation events separated by an interval of 30 ± 10 min, slightly longer than observed in *B. subtilis* (Fig. 4d and Supplementary Fig. 6). Together, these results suggest that the natural process of twin development is similar to that observed in the *B. subtilis* mutants.

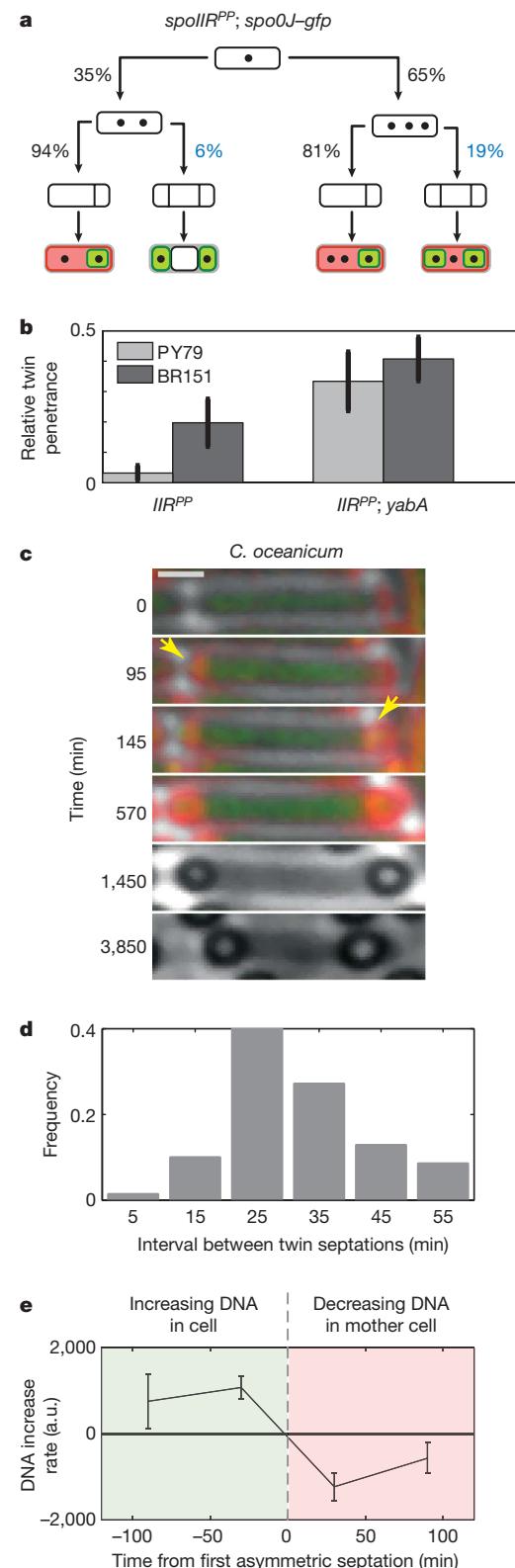


Figure 4 | Evolution of twin sporulation. **a**, Fate tree showing relative frequencies of over-replication (second row) and additional septation (third row) inferred from analysis of terminal fates (bottom row) of $n = 285$ individual cells. Note that the probability of having three compartments depends on chromosome number (blue percentages). Day-to-day variation was $\leq 2\%$ across all measurements. **b**, Strain backgrounds PY79 (used throughout the paper) and BR151 differ in twin penetrance with the same *spoIIR^{PP}* mutation (error bars (s.e.) based on multiple experiments). This difference is reduced by *yabA* mutations. **c–e**, Twin sporulation in *C. oceanicum* resembles that in *B. subtilis* mutants. **c**, Filmstrip showing typical events during *C. oceanicum* sporulation (times in minutes from first frame). Shown are DNA (green), membrane staining (red) and phase contrast (grey). Yellow arrows mark the first appearance of asymmetric septa. **d**, The distribution of time intervals between two septation events during twin sporulation ($n = 70$). **e**, The rate of change of DNA staining was quantified in individual cells. Staining increases before septation (green area), consistent with chromosome replication, and decreases after septation (red area), consistent with transport of DNA into forespores. Data were averaged over $n = 30$ cells owing to cell-cell variability (error bars, s.e.m.).

The concept of developmental canalization was introduced to explain the reproducibility of wild-type development and its contrast with developmental variability in mutants^{26,27}. However, it has been difficult to understand how canalization arises at the molecular level in specific genetic networks and, conversely, how its disruption facilitates the evolution of novel developmental programs. In this case, competition among the core processes of septation, replication and signalling is crucial for generating discrete alternative morphologies (Fig. 3f). Mutations that increase the penetrance of twin sporulation and reduce its dependence on noise provide a gradual mechanism for the stabilization of a discrete change in developmental morphology. The ability to combine single-cell, genetic and evolutionary analyses in bacterial developmental systems like *B. subtilis* sporulation may help to identify basic principles underlying the evolution of developmental mechanisms.

METHODS SUMMARY

Strains and conditions. We derived all *B. subtilis* strains, unless otherwise stated, from parental strain PY79 (ref. 28) using standard protocols²⁹. Details of plasmid and strain construction are described in Methods and Supplementary Methods.

We performed sporulation in liquid culture using the exhaustion method in modified Schaeffer's sporulation medium¹². *Clostridium oceanicum* strain ATCC 25647 (ref. 9), was obtained from the American Type Culture Collection. **Experimental procedure.** We placed *B. subtilis* cells on agarose pads containing sporulation-inducing resuspension medium. The time-lapse microscopy protocol we used is described in Methods. Exposure times were minimized to prevent photodamage. For membrane staining, we added FM4-64 (Invitrogen) at a concentration of 0.2–1 µg ml⁻¹ to the agarose pad before adding cells, or added MitoTracker Green (Invitrogen) at a concentration of 5 µg ml⁻¹. Regulated promoters of the strains spoIIIR^{hyp} (Supplementary Fig. 2) and spoIIE^{hyp} were controlled by addition of appropriate levels of IPTG to the agarose pad. *Clostridium oceanicum* microscopy was similar except anaerobic conditions were maintained using a custom nitrogen flow chamber. Chromosomes were stained with Vybrant DyeCycle Green (Invitrogen).

Quantitative analysis. Quantitative movie analysis used custom image analysis code in MATLAB 7 (Mathworks), similar to that previously described³⁰. Briefly, we segmented phase contrast or fluorescence images using edge detection to identify individual cells. Segmented cells were tracked semi-automatically from frame to frame on the basis of position and orientation. Fluorescence was defined as the sum of pixel intensities within the area of the cell. When the same structure appeared at both poles of the cell (for example in abortively disporic cells), we calculated fluorescence separately for the two halves of the cell. Fate frequencies were manually scored on the basis of criteria detailed in Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 6 December 2008; accepted 15 May 2009.

Published online 5 July; corrected 23 July 2009 (see full-text HTML version for details).

- Horvitz, H. R. & Sulston, J. E. Isolation and genetic characterization of cell-lineage mutants of the nematode *Caenorhabditis elegans*. *Genetics* **96**, 435–454 (1980).
- Queitsch, C., Sangster, T. A. & Lindquist, S. Hsp90 as a capacitor of phenotypic variation. *Nature* **417**, 618–624 (2002).
- Sangster, T. A. et al. HSP90 affects the expression of genetic variation and developmental stability in quantitative traits. *Proc. Natl Acad. Sci. USA* **105**, 2963–2968 (2008).
- Coote, J. G. Sporulation in *Bacillus subtilis*. Characterization of oligosporogenous mutants and comparison of their phenotypes with those of asporogenous mutants. *J. Gen. Microbiol.* **71**, 1–15 (1972).
- Rutherford, S. L. & Henikoff, S. Quantitative epigenetics. *Nature Genet.* **33**, 6–8 (2003).
- Felix, M. A. & Wagner, A. Robustness and evolution: concepts, insights and challenges from a developmental model system. *Heredity* **100**, 132–140 (2008).
- West-Eberhard, M. J. Developmental plasticity and the origin of species differences. *Proc. Natl Acad. Sci. USA* **102** (suppl. 1), 6543–6549 (2005).
- Kirschner, M. & Gerhart, J. *The Plausibility of Life: Resolving Darwin's Dilemma* 71–108 (Yale Univ. Press, 2005).
- Smith, L. D. *Clostridium oceanicum*, sp. n., a sporeforming anaerobe isolated from marine sediments. *J. Bacteriol.* **103**, 811–813 (1970).

- Angert, E. R. Alternatives to binary fission in bacteria. *Nature Rev. Microbiol.* **3**, 214–224 (2005).
- Hilbert, D. W. & Piggot, P. J. Compartmentalization of gene expression during *Bacillus subtilis* spore formation. *Microbiol. Mol. Biol. Rev.* **68**, 234–262 (2004).
- Khvorova, A., Chary, V. K., Hilbert, D. W. & Piggot, P. J. The chromosomal location of the *Bacillus subtilis* sporulation gene *spoIIR* is important for its function. *J. Bacteriol.* **182**, 4425–4429 (2000).
- Zupancic, M. L., Tran, H. & Hofmeister, A. E. Chromosomal organization governs the timing of cell type-specific gene expression required for spore formation in *Bacillus subtilis*. *Mol. Microbiol.* **39**, 1471–1481 (2001).
- Piggot, P. J. & Coote, J. G. Genetic aspects of bacterial endospore formation. *Bacteriol. Rev.* **40**, 908–962 (1976).
- Karow, M. L., Glaser, P. & Piggot, P. J. Identification of a gene, *spoIIR*, that links the activation of sigma E to the transcriptional activity of sigma F during sporulation in *Bacillus subtilis*. *Proc. Natl Acad. Sci. USA* **92**, 2012–2016 (1995).
- Eichenberger, P., Fawcett, P. & Losick, R. A three-protein inhibitor of polar septation during sporulation in *Bacillus subtilis*. *Mol. Microbiol.* **42**, 1147–1162 (2001).
- Pogliano, J. et al. A vital stain for studying membrane dynamics in bacteria: a novel mechanism controlling septation during *Bacillus subtilis* sporulation. *Mol. Microbiol.* **31**, 1149–1159 (1999).
- Dworkin, J. & Losick, R. Developmental commitment in a bacterium. *Cell* **121**, 401–409 (2005).
- Paredes, C. J., Alsaker, K. V. & Papoutsakis, E. T. A comparative genomic view of clostridial sporulation and physiology. *Nature Rev. Microbiol.* **3**, 969–978 (2005).
- Angert, E. R. & Losick, R. M. Propagation by sporulation in the guinea pig symbiont *Metabacterium polyspora*. *Proc. Natl Acad. Sci. USA* **95**, 10218–10223 (1998).
- Dworkin, J. & Losick, R. Does RNA polymerase help drive chromosome segregation in bacteria? *Proc. Natl Acad. Sci. USA* **99**, 14089–14094 (2002).
- Lee, P. S., Lin, D. C., Moriya, S. & Grossman, A. D. Effects of the chromosome partitioning protein SpoOJ (ParB) on oriC positioning and replication initiation in *Bacillus subtilis*. *J. Bacteriol.* **185**, 1326–1337 (2003).
- Ben-Yehuda, S. & Losick, R. Asymmetric cell division in *B. subtilis* involves a spiral-like intermediate of the cytoskeletal protein FtsZ. *Cell* **109**, 257–266 (2002).
- Noirot-Gros, M. F. et al. Functional dissection of YabA, a negative regulator of DNA replication initiation in *Bacillus subtilis*. *Proc. Natl Acad. Sci. USA* **103**, 2368–2373 (2006).
- Keis, S., Shaheen, R. & Jones, D. T. Emended descriptions of *Clostridium acetobutylicum* and *Clostridium beijerinckii*, and descriptions of *Clostridium saccharoperbutylacetonicum* sp. nov. and *Clostridium saccharoacetylum* sp. nov. *Int. J. Syst. Evol. Microbiol.* **51**, 2095–2103 (2001).
- Waddington, C. H. Canalization of development and the inheritance of acquired characters. *Nature* **150**, 563–565 (1942).
- Rutherford, S., Hirate, Y. & Swalla, B. J. The Hsp90 capacitor, developmental remodeling, and evolution: the robustness of gene networks and the curious evolvability of metamorphosis. *Crit. Rev. Biochem. Mol. Biol.* **42**, 355–372 (2007).
- Youngman, P., Perkins, J. B. & Losick, R. Construction of a cloning site near one end of Tn917 into which foreign DNA may be inserted without affecting transposition in *Bacillus subtilis* or expression of the transposon-borne erm gene. *Plasmid* **12**, 1–9 (1984).
- Harwood, C. R. & Cutting, S. M. *Molecular Biological Methods for Bacillus* (Wiley, 1990).
- Rosenfeld, N., Young, J. W., Alon, U., Swain, P. S. & Elowitz, M. B. Gene regulation at the single-cell level. *Science* **307**, 1962–1965 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Leadbetter and E. Matson for their help with the anaerobic species. We thank R. Losick, A. Grossman, M. Fujita and A. Arkin for strains and advice. We thank R. Kishony, D. Jones, Wolfgang Schwarz, G. Suel, J.-G. Ojalvo, B. Shraiman, J. Levine, J. C. W. Locke, D. Sprinzak, L. Cai and other members of M.B.E. and P.J.P.'s labs for helpful discussions. Work in the P.J.P.'s lab was supported by Public Health Service Grant GM43577 from the US National Institutes of Health (NIH). Work in M.B.E.'s lab was supported by NIH grants R01GM079771 and P50 GM068763, US National Science Foundation CAREER Award 0644463 and the Packard Foundation. A.E. was supported by the International Human Frontier Science Organization and the European Molecular Biology Organization.

Author Contributions A.E., V.K.C., J.D., P.J.P. and M.B.E. designed the research; A.E., V.K.C., P.X., M.E.F. and O.C.L. performed the experiments; A.E. and V.K.C. analysed the results; and A.E. and M.B.E. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.B.E. (melowitz@caltech.edu).

METHODS

Construction of LacI-regulated promoters. Briefly, a *lacO* site was added to the *spoIIIR*, *spoIIQ* and *spoIIIE* promoter sequences two base pairs downstream of their -10 RNA polymerase binding sites. Expression was measured and compared with the wild-type promoter using an *mCherry* reporter. The normalized mean time course of gene expression from each of these modified promoters was similar to the corresponding wild-type expression profile and independent of IPTG. See Supplementary Figs 3 and 11 and Supplementary Methods for further details.

Time-lapse microscopy. Cells were grown overnight in CH medium²⁹, diluted back in the morning and grown to an optical density at 600 nm (OD600) of 0.6 at 37 °C. They were then resuspended in sporulation-inducing resuspension medium²⁹ for 1.5 h. Subsequently they were diluted 1:2 and 1 µl of the resulting suspension was placed on a ~0.5-cm-square agarose pad made of 1.5% low-melting-point agarose in resuspension medium (Sigma). Pads were sealed in a glass-bottom dish (Willco). Typically ~10% of cells will commit to sporulation immediately, and the rest will grow for three or more generations on the pad before sporulating asynchronously. Time-lapse images were acquired using an epifluorescence inverted microscope (Olympus IX81) with an automated stage (ASI) and a sensitive camera (Hamamatsu ORCA ER), all coordinated using custom software³⁰. The temperature was fixed at 37 °C using a temperature-controlled environmental chamber.

Quantitative analysis. Quantitative movie analysis was done using custom image analysis code in MATLAB 7 (Mathworks), similar to that previously described^{30–32}. To reduce segmentation errors and leaking of signal from a strongly expressing cell to a neighbour with low expression, fluorescence images were de-blurred using the iterative Lucy–Richardson method (via the MATLAB function ‘deconvlucy’). The resulting expression curves were fit to a piecewise-linear function composed of pre- and post-expression initiation line segments. The results were used to calculate the time of onset of expression (where the two lines meet), the expression level (the difference in slope of the two lines) and other characteristics of the expression profile.

Fate frequencies. To measure the relative frequencies of different primary fates, time-lapse movies were acquired in which sporulation events were marked either by a forespore reporter or by a membrane marker. Sporulating cells were defined as cells that formed a phase-bright spore. Cases in which the forespore failed at a later stage, turning dark again, were also included in the statistics. The majority of sporulating cells continued developing, to produce a mature spore. Abortively dispersive cells were defined as cells that formed a second forespore without mother-cell growth. Most escaping cells elongated as described in the text. A small fraction of cells did not continue to grow and eventually lysed. Another small fraction, classified as escape cells, made two septa in a pattern similar to dispersive cells but mother-cell growth continued in a manner resembling other escaped cells³³. We defined twins as cells in which two forespores were engulfed in one mother cell and continued to produce two phase-bright bodies. We

included cases in which one or both twin spores failed at a later stage, becoming phase dark again. Notably, a large fraction of twins did manage to form two mature spores. We checked 150–500 cells in each experiment in which frequency was measured.

Comparison between strain background and environment. Strains PY79 and BR151 are both derived from the transformable *B. subtilis* 168 strain, but were maintained in different laboratories for several decades³⁴. PY79 was obtained from R. Losick (Harvard University); BR151 was obtained from F. Young (then at the University of Rochester) and maintained in the Piggot lab. For the liquid assay, cells were inoculated from an overnight culture grown at 37 °C in modified Schaeffer's sporulation medium. We defined T_0 as the time at which exponential growth ended. Cells were observed using phase-contrast microscopy 7, 8 and 9 h after T_0 , and their twin/monospore ratio was defined at the time point, giving the maximum value. Errors reflect day-to-day variation.

Clostridium oceanicum general procedures. Cells were saved as frozen spore suspensions and were inoculated with a heat-kill step to additionally select against sporulation mutants. Cells were grown at 37 °C in BHI medium (BD) with 0.5 g l⁻¹ L-cysteine·HCl, and prepared in anaerobic conditions under nitrogen flow. Typically, cells grew exponentially to an optical density at 600 nm of ~1. They then shifted to a transition state for ~7 h before initiating sporulation.

Clostridium oceanicum time-lapse microscopy and analysis. Preconditioned medium was prepared from cultures at the end of the transition stage. It was then mixed with 1.5% low-melting-point agarose (Sigma) and solidified. We used 2 µg ml⁻¹ FM4-64 (Invitrogen) to stain membranes and 50 nM Vybrant DyeCycle Green (Invitrogen) to stain DNA. Cells from the same stage as the media were then placed on the pad. Cells typically overgrew and areas with a monolayer of cells were followed. All cell and media handling was done in an anaerobic glove box (Coy). Samples were moved in a sealed jar from the anaerobic glove box to the microscope. Spores typically matured after 4 d. The data were analysed for the timing of septum formation (as judged by eye from the FM4-64 channel) and for the rate of change of DNA level. For that purpose, individual cells were manually segmented and tracked. DNA-marker fluorescence was corrected for background and summed over the area of the cell (before septation) or the mother cell (after septation).

31. Suel, G. M., Garcia-Ojalvo, J., Liberman, L. M. & Elowitz, M. B. An excitable gene regulatory circuit induces transient cellular differentiation. *Nature* **440**, 545–550 (2006).
32. Suel, G. M., Kulkarni, R. P., Dworkin, J., Garcia-Ojalvo, J. & Elowitz, M. B. Tunability and noise dependence in differentiation dynamics. *Science* **315**, 1716–1719 (2007).
33. Becker, E. C. & Poglino, K. Cell-specific *SpoIIIE* assembly and DNA translocation polarity are dictated by chromosome orientation. *Mol. Microbiol.* **66**, 1066–1079 (2007).
34. Zeigler, D. R. et al. The origins of 168, W23, and other *Bacillus subtilis* legacy strains. *J. Bacteriol.* **190**, 6983–6995 (2008).

LETTERS

Increased mortality and AIDS-like immunopathology in wild chimpanzees infected with SIVcpz

Brandon F. Keele^{1†}, James Holland Jones⁴, Karen A. Terio⁵, Jacob D. Estes⁶, Rebecca S. Rudicell², Michael L. Wilson^{7,8}, Yingying Li¹, Gerald H. Learn¹, T. Mark Beasley³, Joann Schumacher-Stankey⁸, Emily Wroblewski⁸, Anna Mosser⁹, Jane Raphael⁹, Shadrack Kamenya⁹, Elizabeth V. Lonsdorf¹⁰, Dominic A. Travis¹¹, Titus Mlengeya¹², Michael J. Kinsel⁵, James G. Else¹³, Guido Silvestri¹⁴, Jane Goodall¹⁵, Paul M. Sharp¹⁶, George M. Shaw¹, Anne E. Pusey⁸ & Beatrice H. Hahn^{1,2}

African primates are naturally infected with over 40 different simian immunodeficiency viruses (SIVs), two of which have crossed the species barrier and generated human immunodeficiency virus types 1 and 2 (HIV-1 and HIV-2)^{1,2}. Unlike the human viruses, however, SIVs do not generally cause acquired immunodeficiency syndrome (AIDS) in their natural hosts³. Here we show that SIVcpz, the immediate precursor of HIV-1, is pathogenic in free-ranging chimpanzees. By following 94 members of two habituated chimpanzee communities in Gombe National Park, Tanzania, for over 9 years, we found a 10- to 16-fold higher age-corrected death hazard for SIVcpz-infected ($n = 17$) compared to uninfected ($n = 77$) chimpanzees. We also found that SIVcpz-infected females were less likely to give birth and had a higher infant mortality rate than uninfected females. Immunohistochemistry and *in situ* hybridization of post-mortem spleen and lymph node samples from three infected and two uninfected chimpanzees revealed significant CD4⁺ T-cell depletion in all infected individuals, with evidence of high viral replication and extensive follicular dendritic cell virus trapping in one of them. One female, who died within 3 years of acquiring SIVcpz, had histopathological findings consistent with end-stage AIDS. These results indicate that SIVcpz, like HIV-1, is associated with progressive CD4⁺ T-cell loss, lymphatic tissue destruction and premature death. These findings challenge the prevailing view that all natural SIV infections are non-pathogenic and suggest that SIVcpz has a substantial negative impact on the health, reproduction and lifespan of chimpanzees in the wild.

Little is known about the *in vivo* pathogenicity of SIVcpz because, until recently, it has been impossible to identify and monitor infected apes in the wild². Since the first description of SIVcpz in wild-caught chimpanzees, only seven naturally infected apes have been studied in captivity, five of whom died as infants of unknown causes shortly after they were placed in sanctuaries². Only one naturally infected chimpanzee was subjected to virological and immunological analyses, and in this ape, SIVcpz infection was not associated with CD4⁺ T-cell decline, loss of T-cell function, or degenerative changes in lymph node architecture^{4–6}. Similar findings were also reported for

two other chimpanzees inoculated with SIVcpz in captivity⁶. On the basis of these data, it has been assumed that SIVcpz resembles other natural SIV infections of sooty mangabeys and African green monkeys in its non-pathogenic phenotype. However, field and molecular studies suggested otherwise^{7,8}. Epidemiological surveys showed that SIVcpz is less prevalent and much less evenly distributed among wild communities than are SIVs infecting sooty mangabeys and African green monkeys^{2,7,8}. Moreover, SIVcpz was found to be one of very few SIVs to have lost a highly conserved Nef function (that is, the ability to down-modulate the T-cell receptor from the surface of infected CD4⁺ T cells) that correlates with CD4⁺ T-cell preservation in naturally infected primates^{9,10}. Finally, evolutionary analyses revealed that chimpanzees, like humans, acquired SIVcpz relatively more recently by cross-species transmission of SIVs through infected monkeys on which chimpanzees prey¹¹. Collectively these data suggested that the natural history of SIVcpz infection differs from that of other primate lentiviruses. To examine this, we initiated a prospective study in Gombe National Park, Tanzania, the only field site where SIVcpz can be studied in wild-living, yet habituated, chimpanzee communities^{2,7}.

Gombe National Park is located on the shores of Lake Tanganyika and is home to three chimpanzee communities, termed Kasekela (~65 members), Mitumba (~25 members) and Kalande (10–20 members) (Supplementary Fig. 1). The Kasekela and Mitumba chimpanzees have been under continuous observation since the 1960s and 1980s, respectively, and their demography, social structure, reproductive behaviour and individual life histories are well known^{12,13}. The Kalande chimpanzees are not habituated and thus much less well studied. Although the three communities have distinct ranges, interactions between members occur in the form of territorial fights and the migration of adolescent females who typically leave their natal group before having their first offspring. Owing to extensive habitat destruction surrounding the park (Supplementary Fig. 1a), the Gombe chimpanzees have become isolated from other east African ape communities in recent decades. SIVcpz infection has been documented in all three Gombe communities⁷, but only the Mitumba and Kasekela chimpanzees have been studied systematically.

¹Department of Medicine, ²Department of Microbiology, ³Department of Biostatistics, University of Alabama at Birmingham, Birmingham, Alabama 35294, USA. ⁴Department of Anthropology, Stanford University, Stanford, California 94305, USA. ⁵University of Illinois Zoological Pathology Program, Maywood, Illinois 60153, USA. ⁶The AIDS and Cancer Virus Program, Science Applications International Corporation-Frederick Inc., National Cancer Institute-Frederick, Frederick, Maryland 21702, USA. ⁷Department of Anthropology, University of Minnesota, Minneapolis, Minnesota 55455, USA. ⁸Jane Goodall Institute's Center for Primate Studies, Department of Ecology, Evolution and Behavior, University of Minnesota, St Paul, Minnesota 55108, USA. ⁹Gombe Stream Research Centre, The Jane Goodall Institute, Kigoma, Tanzania. ¹⁰The Lester E. Fisher Center for the Study and Conservation of Apes and ¹¹Department of Conservation and Science, Lincoln Park Zoo, Chicago, Illinois 60614, USA. ¹²Tanzania National Parks, Arusha, Tanzania. ¹³Division of Animal Resources, Yerkes National Primate Research Center, Emory University, Atlanta, Georgia 30322, USA. ¹⁴Department of Pathology and Laboratory Medicine, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19107, USA. ¹⁵The Jane Goodall Institute, Arlington, Virginia 22203, USA. ¹⁶Institute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 3JT, UK. [†]Present address: The AIDS and Cancer Virus Program, Science Applications International Corporation-Frederick Inc., National Cancer Institute-Frederick, Frederick, Maryland 21702, USA.

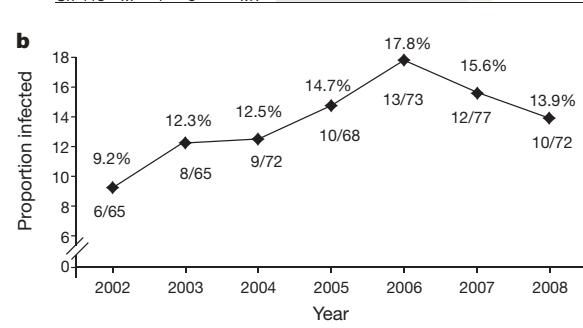
SIVcpz testing of Gombe chimpanzees began in 2000 when non-invasive (faecal- and urine-based) assays for virus-specific antibody and nucleic acid detection were first developed^{2,7}. Since then, 226 urine and 1,153 faecal samples have been collected from 100 chimpanzees (median 11 samples per individual; range 1–70), including 69 from Kasekela and 25 from Mitumba (Supplementary Table 1). Most chimpanzees were sampled at least once every year, and 32 individuals were followed continuously over 7 years. Urine and faecal samples were examined for SIVcpz antibodies by western blot analysis, and a subset of antibody-positive faecal samples was tested for viral RNA by reverse transcription polymerase chain reaction (RT-PCR) amplification^{7,8}. In addition, all chimpanzees were genotyped by amplifying mitochondrial and microsatellite markers from faecal DNA to verify sample identity⁸. Using these non-invasive approaches, we found 13 Kasekela and 4 Mitumba chimpanzees to be antibody-positive, 14 of whom also had detectable SIVcpz RNA in their faeces (Fig. 1). Nine chimpanzees were infected at first analysis, whereas eight others were identified during the course of the study (Fig. 1a). The latter included two infants of SIVcpz-infected mothers (Ch-062 and Ch-103), as well as six individuals who acquired incident infections after exposure to infected community members (Ch-004, Ch-021, Ch-036, Ch-039, Ch-048 and Ch-052). Figure 1b depicts the proportion of SIVcpz-positive chimpanzees in Mitumba and Kasekela over a 7-year observation period. During this time, SIVcpz prevalence rates ranged between 9% and 18%.

To examine the relationship of the newly identified viruses to each other and to previously characterized SIVcpz strains, we constructed phylogenetic trees of available *pol* (481–854 bp) and *env-nef* (279–609 bp) sequences (Supplementary Fig. 2). For these analyses, we also included five SIVcpz strains from the non-habituated Kalande community. The results showed that all Gombe viruses formed a monophyletic lineage within the SIVcpz_{Pts} (SIVcpz from *Pan troglodytes schweinfurthii*) radiation. Within this lineage, viruses from Mitumba, Kasekela and Kalande were interspersed, indicating frequent transfer of SIVcpz between communities, most probably via migrating females. Phylogenetic analyses also identified clusters of epidemiologically linked infections, indicating horizontal transmission. One such cluster included four Kasekela chimpanzees (Ch-004, Ch-021, Ch-036 and Ch-052), all of whom became infected within a 20 month time period (Supplementary Table 2). In humans, primary HIV-1 infection is responsible for a substantial portion of new transmissions because acutely infected individuals have higher viral loads and are more infectious than chronically infected individuals¹⁴. The finding of nearly identical viruses in four chimpanzees thus suggests that some of them acquired SIVcpz from acutely infected mating partners. Closely related viruses were also observed in a presumed mother/daughter pair (Ch-071 and Ch-099), suggesting vertical transmission, but this conclusion remains tentative because both females were already sexually active at the time of first sampling (Supplementary Table 3). In addition, vertical transmission is suspected in two infants (Ch-062 and Ch-103) who were faecal-antibody-positive (but viral RNA-negative) at the age of 1.8 and 2.6 years, respectively. Interestingly, one of these (Ch-103) was born to a SIVcpz-negative mother (Ch-021) who became infected 10–15 months after giving birth (Supplementary Table 3). Thus, the infant seems to have acquired SIVcpz by breast milk transmission.

Together, these data suggest that SIVcpz, like HIV-1, is transmitted by both horizontal and vertical routes.

To determine whether SIVcpz infection has an impact on chimpanzee survival, we used discrete event-history methods to model the

Chimpanzee ID	Sex	Faecal samples	Urine samples	Community									Death
				2000	2001	2002	2003	2004	2005	2006	2007	2008	
Ch-001	F	23	10	KK	0	0	0	0	0	0	0	0	DLS 23 Jan 07
Ch-002	F	12	0	KK	0	0	0	0	0	0	0	0	DLS 6 Nov 02
Ch-003	M	18	2	KK	0	0	0	0	0	0	0	0	DLS 1 Sep 04
Ch-004	M	21	3	KK	1	1	1	1	1	1	1	1	DLS 24 Aug 04
Ch-005	M	20	3	KK	0	0	0	0	0	0	0	0	DOD 3 Oct 05
Ch-006	M	15	2	KK	0	0	0	0	0	0	0	0	DOD 5 Nov 06
Ch-007	M	10	4	KK	0	0	0	0	0	0	0	0	DOD 8 Dec 03
Ch-008	M	5	5	KK	0	0	0	0	0	0	0	0	DOD 2 Oct 07
Ch-009	F	19	13	KK	0	0	0	0	0	0	0	0	DLS 25 Jun 03
Ch-010	M	11	5	KK	0	0	0	0	0	0	0	0	DOD 10 Sep 01
Ch-011	F	15	11	KK	0	0	0	0	0	0	0	0	DOD 17 Jan 05
Ch-012	F	12	8	KK	0	0	0	0	0	0	0	0	DOD 13 May 08
Ch-013	M	18	4	KK	0	0	0	0	0	0	0	0	DOD 14 Jul 04
Ch-014	M	12	1	KK	0	0	0	0	0	0	0	0	DOD 12 Jul 07
Ch-015	F	17	12	KK	0	0	0	0	0	0	0	0	DOD 13 Jun 04
Ch-016	M	14	9	KK	0	0	0	0	0	0	0	0	DOD 18 Nov 06
Ch-017	F	12	3	KK	0	0	0	0	0	0	0	0	DOD 19 Jul 04
Ch-018	F	10	1	KK	0	0	0	0	0	0	0	0	DOD 13 Jul 05
Ch-019	F	11	1	KK	0	0	0	0	0	0	0	0	DOD 13 Jul 05
Ch-020	M	15	3	KK	0	0	0	0	0	0	0	0	DOD 13 Jul 05
Ch-021	F	11	2	(KL) KK	1	1	1	1	1	1	1	1	DOD 13 Jul 05
Ch-022	F	17	2	(KL) KK	1	1	1	1	1	1	1	1	DOD 13 Jul 05
Ch-023	M	15	2	KK	0	0	0	0	0	0	0	0	DOD 13 Jul 05
Ch-024	F	9	10	KK	0	0	0	0	0	0	0	0	DOD 13 Jul 05
Ch-025	F	16	11	KK	0	0	0	0	0	0	0	0	DOD 13 Jul 05
Ch-026	F	11	6	KK	0	0	0	0	0	0	0	0	DOD 13 Jul 05
Ch-027	M	17	5	KK	0	0	0	0	0	0	0	0	DOD 13 Jul 05
Ch-028	F	9	6	KK	0	0	0	0	0	0	0	0	DOD 13 Jul 05
Ch-029	F	11	7	(KL) KK	0	0	0	0	0	0	0	0	DOD 13 Jul 05
Ch-030	F	5	4	KK	1	1	1	1	1	1	1	1	DOD 13 Jul 05
Ch-031	F	27	10	KK	0	0	0	0	0	0	0	0	DOD 13 Jul 05
Ch-032	F	12	8	KK	0	0	0	0	0	0	0	0	DOD 13 Jul 05
Ch-033	F	8	1	KK	1	1	1	1	1	1	1	1	DOD 13 Jul 05
Ch-034	M	10	4	KK	0	0	0	0	0	0	0	0	DOD 13 Jul 05
Ch-035	M	15	3	KK	0	0	0	0	0	0	0	0	DOD 13 Jul 05
Ch-036	F	7	4	KK	0	0	0	0	0	0	0	0	DOD 13 Jul 05
Ch-041	F	10	1	KK	0	0	0	0	0	0	0	0	DOD 13 Jul 05
Ch-050	M	11	3	KK	0	0	0	0	0	0	0	0	DOD 13 Jul 05
Ch-051	M	12	4	KK	0	0	0	0	0	0	0	0	DOD 13 Jul 05
Ch-052	M	8	0	KK	0	0	0	0	0	0	0	0	DOD 13 Jul 05
Ch-053	F	13	0	KK	0	0	0	0	0	0	0	0	DOD 13 Jul 05
Ch-054	M	10	5	KK	0	0	0	0	0	0	0	0	DOD 13 Jul 05
Ch-055	F	11	0	KK	0	0	0	0	0	0	0	0	DOD 13 Jul 05
Ch-056	M	6	3	KK	0	0	0	0	0	0	0	0	DOD 13 Jul 05
Ch-057	F	15	7	KK	0	0	0	0	0	0	0	0	DOD 13 Jul 05
Ch-058	M	13	2	KK	0	0	0	0	0	0	0	0	DOD 13 Jul 05
Ch-060	M	11	0	KK	0	0	0	0	0	0	0	0	DOD 13 Jul 05
Ch-061	F	9	0	KK	0	0	0	0	0	0	0	0	DOD 13 Jul 05
Ch-062	M	1	0	KK	1	1	1	1	1	1	1	1	DOD 13 Jun 04
Ch-071	F	10	0	(KL) KK	1	1	1	1	1	1	1	1	DOD 18 Nov 06
Ch-072	M	1	0	KK	0	0	0	0	0	0	0	0	DOD 18 Nov 06
Ch-073	F	7	0	KK	0	0	0	0	0	0	0	0	DOD 18 Nov 06
Ch-074	M	4	0	KK	0	0	0	0	0	0	0	0	DOD 18 Nov 06
Ch-078	F	25	0	(MT) KK	0	0	0	0	0	0	0	0	DOD 22 Dec 04
Ch-079	F	16	3	(KL) KK	1	1	1	1	1	1	1	1	DOD 22 Dec 04
Ch-080	F	14	0	(MT) KK	1	1	1	1	1	1	1	1	DOD 22 Dec 04
Ch-090	M	2	0	KK	0	0	0	0	0	0	0	0	DOD 13 Nov 06
Ch-097	F	5	0	KK	0	0	0	0	0	0	0	0	DOD 13 Nov 06
Ch-099	F	11	0	(KL) KK	1	1	1	1	1	1	1	1	DOD 13 Nov 06
Ch-101	F	6	0	(KL) KK	0	0	0	0	0	0	0	0	DOD 13 Nov 06
Ch-102	F	7	0	KK	0	0	0	0	0	0	0	0	DOD 13 Nov 06
Ch-103	F	2	0	KK	0	0	0	0	0	0	0	0	DOD 13 Nov 06
Ch-104	F	3	0	KK	0	0	0	0	0	0	0	0	DOD 2 Oct 07
Ch-105	F	5	0	(KL) KK	0	0	0	0	0	0	0	0	DOD 13 May 08
Ch-108	F	2	0	(KL) KK	0	0	0	0	0	0	0	0	DOD 13 May 08
Ch-114	F	2	0	KK	0	0	0	0	0	0	0	0	DOD 13 May 08
Ch-115	M	3	0	KK	0	0	0	0	0	0	0	0	DOD 13 May 08
Ch-116	M	1	0	KK	0	0	0	0	0	0	0	0	DOD 13 May 08
Ch-117	F	1	0	KK	0	0	0	0	0	0	0	0	DOD 13 May 08
Ch-037	M	0	4	MT	1	0	0	0	0	0	0	0	DOD 10 Sep 01
Ch-038	F	3	1	MT	0	0	0	0	0	0	0	0	DOD 25 Jun 03
Ch-039	F	25	0	MT	0	0	0	0	0	0	0	0	DOD 22 Dec 04
Ch-042	F	21	0	MT	0	0	0	0	0	0	0	0	DOD 22 Dec 04
Ch-044	F	24	0	MT	0	0	0	0	0	0	0	0	DOD 22 Dec 04
Ch-045	M	21	0	MT	1	1	1	1	1	1	1	1	DOD 22 Dec 04
Ch-046	F	20	2	MT	0	0	0	0	0	0	0	0	DOD 22 Dec 04
Ch-047	F	15	1	MT	0	0	0	0	0	0	0	0	DOD 22 Dec 04
Ch-048	F	24	0	MT	0	0	0	0	0	0	0	0	DOD 22 Dec 04
Ch-049	F	13	0	MT	0	0	0	0	0	0	0	0	DOD 22 Dec 04
Ch-059	M	18	0	MT	0	0	0	0	0	0	0	0	DOD 22 Dec 04
Ch-063	M	7	0	MT	0	0	0	0	0	0	0	0	DOD 22 Dec 04
Ch-065	M	17	1	MT	0	0	0	0	0	0	0	0	DOD 22 Dec 04
Ch-066	F	23	4	MT	0	0	0	0	0	0	0	0	DOD 22 Dec 04
Ch-067	M	5	0	MT	0	0	0	0	0	0	0	0	DOD 22 Dec 04
Ch-068	M	13	0	MT	0	0	0	0	0	0	0	0	DOD 22 Dec 04
Ch-069	M	3	0	MT	0	0	0	0	0	0	0	0	DOD 22 Dec 04
Ch-075	M	4	0	MT	0	0	0	0	0	0	0	0	DOD 22 Dec 04
Ch-076	F	5	0	MT	0	0	0	0	0	0	0	0	DOD 22 Dec 04
Ch-094	F	2	0	MT	0	0	0	0	0	0	0	0	DOD 22 Dec 04
Ch-096	F	8	0	MT	0	0	0	0	0	0	0	0	DOD 22 Dec 04
Ch-098	F	19	0	(KL) MT	0	0	0	0	0	0	0	0	DOD 22 Dec 04
Ch-111	F	3	0	MT	0	0	0	0	0	0	0	0	DOD 22 Dec 04
Ch-112	F	2	0	MT	0	0	0	0	0	0	0	0	DOD 22 Dec 04
Ch-113	M	4	0	MT	0	0	0	0	0	0	0	0	DOD 22 Dec 04



hazard of death for both infected and uninfected individuals. For this analysis, only Kasekela and Mitumba chimpanzees of known SIVcpz infection status were included. During the 9-year observation period, 7 of 17 infected and 11 of 77 uninfected chimpanzees died or disappeared. Death of missing individuals was defined based on previously reported criteria^{15,16}, and for all but one chimpanzee for whom only urine samples were available (Ch-037), infection was confirmed using genotyped faecal samples. Age, sex, SIVcpz infection and death status were tabulated for each individual in each year, yielding over 550 chimpanzee years of observation for analysis. Figure 2 compares the survival rates of SIVcpz-infected (red) and uninfected (blue) chimpanzees using maximally and minimally conservative event definitions. For a maximally conservative estimate (Fig. 2a), the two faecal antibody-positive but viral RNA-negative infants were classified as SIVcpz negative, and two uninfected females who went missing for more than 1 year were counted as being dead (this resulted in the removal of one potentially SIVcpz-related death and the addition of two SIVcpz-unrelated deaths). For a less conservative estimate (Fig. 2b), the two infants were assumed to be infected, and the two females were censored, that is, assumed to have migrated to Kalande or outside the park (this resulted in the addition of one SIVcpz-related death and the removal of two SIVcpz-unrelated deaths). The results showed that SIVcpz-infected chimpanzees died at a significantly faster rate, even under the most stringent assumptions. The maximally conservative approach yielded a 9.8-fold increased death hazard for SIVcpz-positive chimpanzees (95% confidence interval = 2.8–34.3; $P < 0.0001$), whereas the less conservative approach indicated a 15.6-fold increased hazard (95% confidence interval = 4.7–51.8; $P < 0.0001$).

We also examined the impact of SIVcpz infection on female fertility and infant survival. During the 9-year study period, 22 of 30 uninfected females gave birth to 30 infants, whereas 4 of 9 SIVcpz-infected females gave birth to 4 infants (only Kasekela and Mitumba females of known SIVcpz status at the time of delivery were included). For each female, age, SIVcpz status and the risk of becoming pregnant (taking into account adolescent subfecundity, lactational amenorrhea and immigration status; see Methods) were tabulated, and the number of births per total number of risk years was determined. This yielded significantly different ($P = 0.034$) birth rates of 0.32 (29 births per 90 female years) and 0.13 (4 births per 30 female years) for uninfected and infected females, respectively, indicating that the odds of an infected female giving birth were 3.1 times lower than those of an uninfected female. In addition, there was a significant association between the SIVcpz status of the mother and the survival of the infant. Of the 30 babies born to uninfected mothers, only six died before the age of 1 year. In contrast, all four infants born to SIVcpz-infected mothers died before their first birthday (Supplementary Table 4). Therefore, infants born to

SIVcpz-infected mothers had a significantly higher mortality rate than infants born to uninfected mothers ($P = 0.005$).

One Kasekela female (Ch-036), who became infected during the course of the study (between December 2004 and March 2006), died within 3 years of acquiring SIVcpz (November 2007; Fig. 1a). She had no observable injuries, but exhibited profound weakness and lethargy several days before her death. At necropsy, she had severe cachexia with skeletal muscle and hepatocellular atrophy, as well as multiple abdominal abscesses due to nematode infestation. Histologically, pancreatic and mesenteric lymph nodes had marked depletion of both cortical and paracortical lymphocytes (Fig. 3a). Specific staining of spleen sections with B-cell and T-cell markers (CD79a and CD3, respectively) confirmed severe lymphopenia and showed that over 95% of the remaining lymphocytes were B cells (Fig. 3b and c). The spleen also contained multiple areas of follicular hyalinization, which is a hallmark of lymphatic tissue destruction. These histopathological findings are consistent with end-stage HIV-1 and SIVmac infections^{17,18}, and suggested that Ch-036 died of an AIDS-like illness.

To examine whether the lymphatic depletion in Ch-036 represented a rare case of immunodeficiency known to occur sporadically in non-pathogenic SIV infections¹⁹, we obtained post-mortem spleen samples from two additional infected (Ch-045 and Ch-099) and uninfected (Ch-016 and Ch-069) chimpanzees. Three of these (including both infected chimpanzees) died of trauma-related causes, in one instance following intra-group aggression (Ch-045) and in the two others, following spinal cord injury (Ch-069 and Ch-099); the fourth (uninfected) chimpanzee (Ch-016) died of age-related conditions shortly before reaching his 40th birthday.

Spleen sections from these four chimpanzees, as well as from Ch-036, were stained with CD4-specific antibodies (as well as CD3 and CD20 antibodies as controls). CD4⁺ T cells were then quantified by determining the percentage area of periarteriolar lymphoid sheaths (PALS, the T-cell zone equivalent in lymph nodes) that stained positive for CD4. The results revealed that each of the three infected chimpanzees had fewer CD4⁺ T cells in their spleen than either of the two uninfected controls (Fig. 4). These differences were statistically significant whether chimpanzees were compared individually or as groups. The most profound CD4⁺ T-cell loss was seen in Ch-036, whose spleen was also depleted of CD3⁺ T cells and CD20⁺ B cells (Supplementary Fig. 3). A reduction of CD4⁺ T cells was also apparent in the PALS of Ch-045 and Ch-099; however, this depletion was less pronounced, especially in Ch-099, and occurred in the absence of a concurrent loss of other CD3⁺ T-cell and B-cell populations (Supplementary Fig. 3). Thus, in the two infected apes who died of trauma-related causes, CD4⁺ T-cells were selectively depleted whereas other lymphocyte populations remained intact, suggesting that SIVcpz is associated with progressive CD4⁺ T-cell killing through all stages of infection, including clinical latency.

Splenic tissues were also examined for evidence of collagen deposition in the PALS, which is a non-specific marker of chronic immune activation^{20,21}. Although some collagen was detected in the PALS of Ch-099 and the two uninfected chimpanzees, this amount was substantially less than that detected in the PALS of Ch-045, and in particular Ch-036 (Supplementary Fig. 3). In HIV-1-infected humans, collagen deposition in T-cell zones and PALS correlates with CD4⁺ T-cell depletion, altered CD4 T-cell homeostasis and overall immune dysfunction^{20,21}. Consistent with this, fibrotic scarring of the PALS was most severe in Ch-036, intermediate in Ch-045, and minimal in Ch-099 (also see Fig. 4). Thus, SIVcpz seems to cause immune deficiency by mechanisms that are very similar to those of HIV-1. Indeed, *in situ* hybridization using SIVcpz-specific RNA probes revealed numerous productively infected lymphocytes in the spleen of Ch-045, documenting active viral replication and extensive trapping of virus particles by follicular dendritic cells (Supplementary Fig. 4).

We show here that SIVcpz-infected chimpanzees in Gombe have a 10–16-fold increased death hazard compared to uninfected chimpanzees, and that infected females are less likely to give birth and

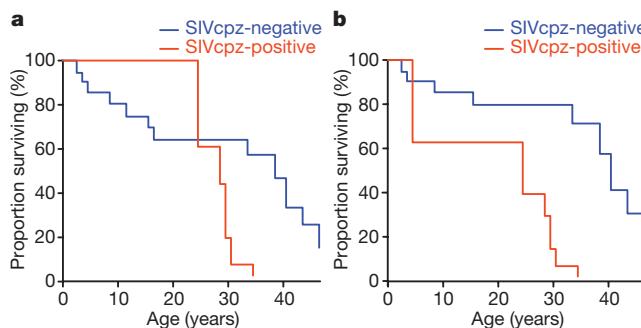


Figure 2 | SIVcpz-associated mortality in Gombe. **a, b**, Kaplan–Meier survival curves are shown for SIVcpz-infected (red) and uninfected (blue) chimpanzees using maximally conservative (**a**) and less conservative (**b**) event definitions (see text for details). Under both scenarios, SIVcpz-infected chimpanzees die at a significantly faster rate (9.8- and 15.6-fold increased death hazard, respectively; $P < 0.0001$).

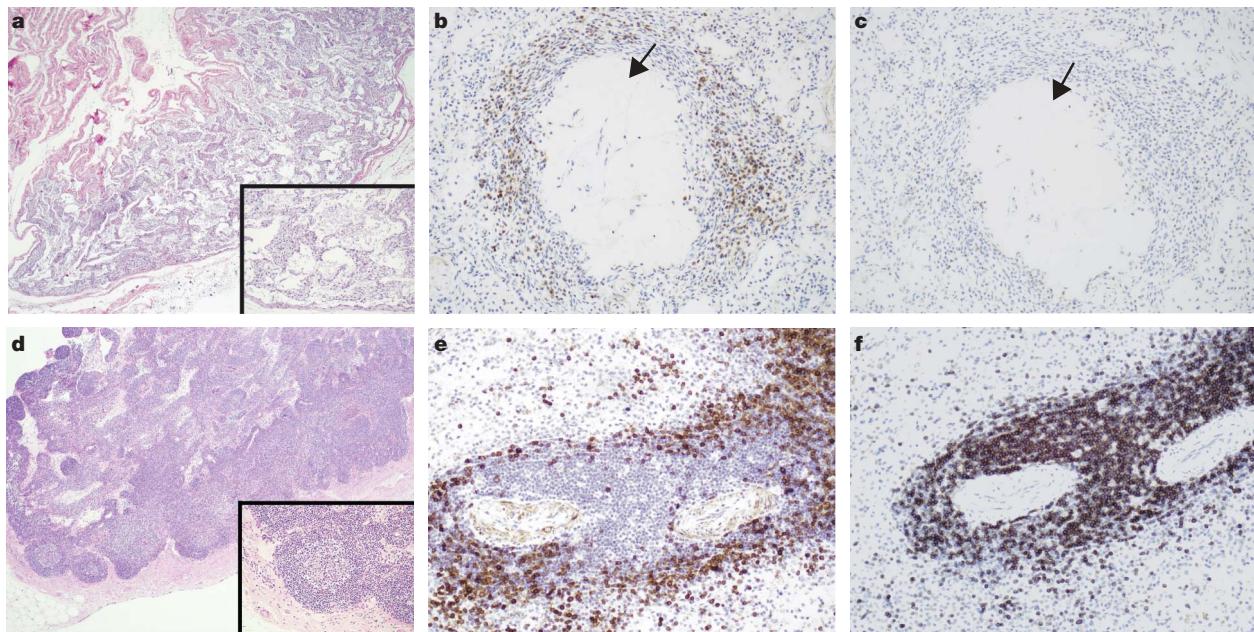


Figure 3 | Histopathology of end-stage SIVcpz infection. **a–c**, Severe lymphatic depletion in mesenteric lymph node (**a**) and splenic tissue (**b**, **c**) of a Kasekela female (Ch-036) who died of an AIDS-like illness within 3 years of acquiring SIVcpz. **d–f**, Normal mesenteric lymph node with multiple primary and secondary lymphoid follicles (**d**) and normal splenic tissue with PALS (**e**, **f**) from an uninfected chimpanzee. **a**, **d**, Lymph node sections

stained with haematoxylin and eosin (nuclei are blue). **b**, **c**, **e**, **f**, Spleen sections stained with anti-CD79a (**b**, **e**) and anti-CD3 (**c**, **f**) antibodies (positive cells are dark brown). Note the severe loss of both B (CD79a) and T (CD3) cells in the PALS of Ch-036. Arrows indicate follicular hyalinization. Original magnification $\times 4$ (**a**, **d**) and $\times 20$ (**b**, **c**, **e**, **f** and insets).

have fewer surviving offspring. We also show that SIVcpz infection is associated with progressive CD4⁺ T-cell loss and immune system destruction, which are hallmarks of pathogenic HIV-1 infection. Although based on a limited sample size, these results suggest that SIVcpz, like HIV-1 in humans^{22–25}, has a substantial negative impact on the health, reproduction and lifespan of wild chimpanzees. What is less certain is the magnitude of this effect and its contribution to the population decline of chimpanzees in Gombe^{16,26} and elsewhere. Before the introduction of highly active antiretroviral therapy, HIV-1-infected humans had a 18–60-fold higher risk of death than uninfected individuals^{23–25}, whereas HIV-2-infected humans had a 2–5-fold higher risk^{25,27}. In contrast, SIVsmm, the immediate precursor of HIV-2 (ref. 1), does not seem to affect the lifespan of

its host, as evidenced by nearly identical survival curves of infected ($n = 167$) and uninfected ($n = 62$) sooty mangabeys at the Yerkes primate centre (Supplementary Fig. 5). On the basis of these mortality data, SIVcpz seems to be less pathogenic than HIV-1, but more pathogenic than HIV-2 and SIVsmm. Although the generality of these results remains to be confirmed in a larger number of chimpanzees the finding that SIVcpz is pathogenic in Gombe provides a unique opportunity to compare disease-causing mechanisms of two closely related viruses in two closely related species. Such an approach is likely to accelerate the identification of viral and host factors responsible for HIV/SIV disease progression, and could lead to novel therapeutic and preventive measures that will benefit both humans and chimpanzees.

METHODS SUMMARY

Observational data. Kasekela and Mitumba chimpanzees have been observed daily since the 1960s and 1980s, respectively, providing a continuous long-term record of community demography and individual behaviour^{12,13,16,26}.

Non-invasive SIVcpz testing. Chimpanzee faecal and urine samples were tested for virus-specific antibodies by western blot analysis, and SIVcpz infection was confirmed by RT-PCR (GenBank accession numbers FJ895381–FJ895405)^{7,8}. All faecal samples used for SIVcpz status determination were genotyped (Supplementary Table 1).

Phylogenetic analyses. Phylogenetic trees of SIVcpz *pol* and *env-nef* sequences were inferred by Bayesian methods²⁸ using a general time-reversible (GTR) likelihood model.

Mortality analyses. SIVcpz-associated mortality hazards were estimated using discrete event history methods as a function of measured covariates (age, sex and SIVcpz status)²⁹. SIVsmm-associated survival differences were examined using a Cox proportional hazards model.

Necropsy. Complete necropsies were performed on five chimpanzees whose bodies were recovered within 6–18 h of death. Pathological and histological examinations were performed on all organs.

Immunohistochemistry, *in situ* hybridization and quantitative image analysis. Post-mortem spleen and lymph node samples were subjected to *in situ* hybridization, immunohistochemical staining and quantitative image analysis as described previously^{20,21}. Splenic CD4⁺ T cells were quantified by manually extracting the images of representative PALS into Photoshop (Adobe), and by determining the percentage of area that stained positive for CD4 (refs 20, 21).

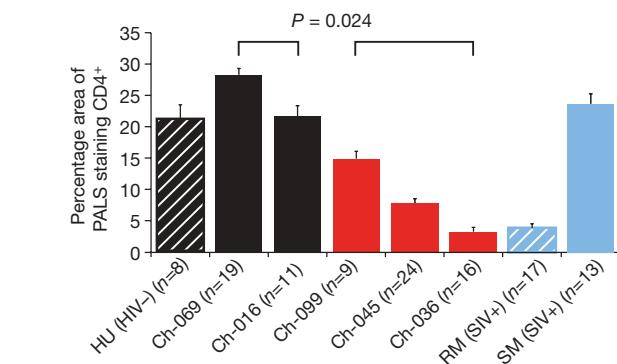


Figure 4 | CD4⁺ T-cell depletion in SIVcpz-infected chimpanzees.

Quantitative image analysis of splenic CD4⁺ T cells in SIVcpz-infected (red bars) and uninfected (solid black bars) chimpanzees, as well as in uninfected human (HU, black hatched bar), SIVmac-infected rhesus macaque (RM, blue hatched bar), and SIVsmm-infected sooty mangabey (SM, solid blue bar) controls. Representative PALS (8–24 per individual) were extracted and the percentage of area that stained positive for CD4 was determined^{20,21}. Average values are shown, with error bars indicating s.e.m. All three SIVcpz-infected chimpanzees had significantly lower CD4⁺ T-cell counts than the two uninfected controls (ANOVA *F*-test; $P = 0.024$). The SIVmac-infected rhesus macaque had end-stage AIDS.

Statistical methods. Splenic CD4⁺ T-cell counts of three infected and two uninfected chimpanzees were compared using ANOVA F-tests. Fertility and infant mortality data for infected and uninfected females were compared using Fisher's exact test.

Received 21 April; accepted 5 June 2009.

- Hahn, B. H., Shaw, G. M., De Cock, K. M. & Sharp, P. M. AIDS as a zoonosis: scientific and public health implications. *Science* **287**, 607–614 (2000).
- Sharp, P. M., Shaw, G. M. & Hahn, B. H. Simian immunodeficiency virus infection of chimpanzees. *J. Virol.* **79**, 3891–3902 (2005).
- Silvestri, G. Immunity in natural SIV infections. *J. Intern. Med.* **265**, 97–109 (2008).
- Koopman, G., Haaksma, A. G., ten Velden, J., Hack, C. E. & Heeney, J. L. The relative resistance of HIV type 1-infected chimpanzees to AIDS correlates with the maintenance of follicular architecture and the absence of infiltration by CD8⁺ cytotoxic T lymphocytes. *AIDS Res. Hum. Retroviruses* **15**, 365–373 (1999).
- Gougeon, M. L. et al. Lack of chronic immune activation in HIV-infected chimpanzees correlates with the resistance of T cells to Fas/Apo-1 (CD95)-induced apoptosis and preservation of a T helper 1 phenotype. *J. Immunol.* **158**, 2964–2976 (1997).
- Heeney, J. L. et al. Transmission of simian immunodeficiency virus SIVcpz and the evolution of infection in the presence and absence of concurrent human immunodeficiency virus type 1 infection in chimpanzees. *J. Virol.* **80**, 7208–7218 (2006).
- Santiago, M. L. et al. Foci of endemic simian immunodeficiency virus infection in wild-living eastern chimpanzees (*Pan troglodytes schweinfurthii*). *J. Virol.* **77**, 7545–7562 (2003).
- Keele, B. F. et al. Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. *Science* **313**, 523–526 (2006).
- Schindler, M. et al. Nef-mediated suppression of T cell activation was lost in a lentiviral lineage that gave rise to HIV-1. *Cell* **125**, 1055–1067 (2006).
- Schindler, M. et al. Inefficient Nef-mediated downmodulation of CD3 and MHC-I correlates with loss of CD4⁺ T cells in natural SIV infection. *PLoS Pathog.* **4**, e1000107 (2008).
- Bailes, E. et al. Hybrid origin of SIV in chimpanzees. *Science* **300**, 1713 (2003).
- Goodall, J. *The Chimpanzees of Gombe: Patterns of Behavior* (Belknap, 1986).
- Pusey, A. E., Pintea, L., Wilson, M. L., Kamenya, S. & Goodall, J. The contribution of long-term research at Gombe National Park to chimpanzee conservation. *Conserv. Biol.* **21**, 623–634 (2007).
- Wawer, M. J. et al. Rates of HIV-1 transmission per coital act, by stage of HIV-1 infection, in Rakai, Uganda. *J. Infect. Dis.* **191**, 1403–1409 (2005).
- Hill, K. et al. Mortality rates among wild chimpanzees. *J. Hum. Evol.* **40**, 437–450 (2001).
- Williams, J. M. et al. Causes of death in the Kasekela chimpanzees of Gombe National Park, Tanzania. *Am. J. Primatol.* **70**, 766–777 (2008).
- Pantaleo, G. et al. HIV infection is active and progressive in lymphoid tissue during the clinically latent stage of disease. *Nature* **362**, 355–358 (1993).
- King, N. W. Jr. Simian immunodeficiency virus infections. In *Nonhuman Primates I* (eds Jones T. C., Mohr, U. & Hunt, R. D.) 5–20 (Springer, 1993).
- Pandrea, I., Silvestri, G. & Apetrei, C. AIDS in African nonhuman primate hosts of SIVs: a new paradigm of SIV infection. *Curr. HIV Res.* **7**, 57–72 (2009).
- Estes, J. D. et al. Early resolution of acute immune activation and induction of PD-1 in SIV-infected sooty mangabeys distinguishes nonpathogenic from pathogenic infection in rhesus macaques. *J. Immunol.* **180**, 6798–6807 (2008).
- Estes, J. D. et al. Simian immunodeficiency virus-induced lymphatic tissue fibrosis is mediated by transforming growth factor β1-positive regulatory T cells and begins in early infection. *J. Infect. Dis.* **195**, 551–561 (2007).
- Brocklehurst, P. & French, R. The association between maternal HIV infection and perinatal outcome: a systematic review of the literature and meta-analysis. *Br. J. Obstet. Gynaecol.* **105**, 836–848 (1998).
- Bhaskaran, K. et al. Changes in the risk of death after HIV seroconversion compared with mortality in the general population. *J. Am. Med. Assoc.* **300**, 51–59 (2008).
- Mulder, D. W. et al. Two-year HIV-1-associated mortality in a Ugandan rural population. *Lancet* **343**, 1021–1023 (1994).
- Hansmann, A. et al. Baseline plasma viral load and CD4 cell percentage predict survival in HIV-1- and HIV-2-infected women in a community-based cohort in The Gambia. *J. Acquir. Immune Defic. Syndr.* **38**, 335–341 (2005).
- Pusey, A. E., Wilson, M. L. & Collins, D. A. Human impacts, disease risk, and population dynamics in the chimpanzees of Gombe National Park, Tanzania. *Am. J. Primatol.* **70**, 738–744 (2008).
- Poulsen, A. G. et al. 9-year HIV-2-associated mortality in an urban community in Bissau, west Africa. *Lancet* **349**, 911–914 (1997).
- Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574 (2003).
- Singer, J. D. & Willett, J. B. *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence* (Oxford Univ. Press, 2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the field staff at the Gombe Stream Research Centre for collecting behavioural data as well as urine and faecal samples from wild chimpanzees; E. Kaaya for help with necropsies; A. Collins for logistical support; the Tanzania Commission for Science and Technology, the Tanzania Wildlife Research Institute, and the Tanzania National Parks for permission to conduct research in Gombe; I. White for discussions; L. Lowenstein for histological consultation; M. Salazar and Y. Chen for technical assistance; and J. C. White for artwork and manuscript preparation. This work was supported by grants from the National Institutes of Health (R01 AI50529, R01 AI58715, U19 AI067854, T32 GM008111), the National Cancer Institute (contract HHSN266200400088C), the UAB Center for AIDS Research (P30 AI 27767), the Yerkes National Primate Research Center (RR-00165), the National Science Foundation (DBS-9021946, SBR-9319909, BSC-0452315, IIS-0431141, BSC-0648481), the Jane Goodall Institute, the Harris Steel Group, the University of Minnesota, the University of Illinois, the US Fish and Wildlife Service Great Ape Conservation Fund, the Windibrow, Arcus, Guthman and Davee Foundations, and the Lincoln Park Zoo. R.S.R. was funded by a Howard Hughes Medical Institute Med-into-Grad Fellowship. The content of this publication does not necessarily reflect the views or policies of the US Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government.

Author Contributions All authors contributed to the acquisition, analysis and interpretation of the data; B.F.K., A.E.P., J.G., G.M.S., P.M.S. and B.H.H. initiated and conceived the study; B.F.K., R.S.R., Y.L. and E.W. performed non-invasive SIVcpz screening and genetic analyses; J.H.J., J.S.-S., M.L.W. and A.E.P. analysed chimpanzee fertility and mortality; K.A.T., M.J.K., J.R., D.A.T. and T.M. performed necropsies; K.A.T. and J.D.E performed immunohistochemistry and *in situ* hybridization studies; G.H.L. and P.M.S. performed phylogenetic analyses; J.H.J., T.M.B. and P.M.S. performed statistical analyses; M.L.W., A.M., S.K., E.V.L. and D.A.T. conducted and supervised all fieldwork. J.G.E. and G.S. provided data on captive sooty mangabeys; and B.F.K., A.E.P., P.M.S., G.M.S. and B.H.H. coordinated the contributions of all authors and wrote the paper. B.F.K., J.H.J., K.A.T., J.D.E., R.S.R. and M.L.W. contributed equally to this paper.

Author Information Newly derived SIVcpz sequences have been deposited in the GenBank database under accession numbers FJ895381–FJ895405. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to B.H.H. (bhahn@uab.edu).

LETTERS

Regulation of the innate immune response by threonine-phosphatase of Eyes absent

Yasutaka Okabe^{1,2†*}, Teruyuki Sano^{1*} & Shigekazu Nagata^{1,2}

Innate immunity is stimulated not only by viral or bacterial components, but also by non-microbial danger signals (damage-associated molecular patterns)¹. One of the damage-associated molecular patterns is chromosomal DNA that escapes degradation. In programmed cell death and erythropoiesis, DNA from dead cells or nuclei expelled from erythroblasts is digested by DNase II in the macrophages after they are engulfed. *DNase II*^{-/-} (also known as *Dnase2a*^{-/-}) mice suffer from severe anaemia or chronic arthritis due to interferon- β (IFN- β) and tumour necrosis factor- α (TNF- α) produced from the macrophages carrying undigested DNA^{2,3} in a Toll-like receptor (TLR)-independent mechanism⁴. Here we show that Eyes absent 4 (EYA4), originally identified as a co-transcription factor, stimulates the expression of IFN- β and CXCL10 in response to the undigested DNA of apoptotic cells. EYA4 enhanced the innate immune response against viruses (Newcastle disease virus and vesicular stomatitis virus), and could associate with signalling molecules (IPS-1 (also known as MAVS), STING (TMEM173) and NLRX1). Three groups have previously shown that EYA has phosphatase activity⁵⁻⁷. We found that mouse EYA family members act as a phosphatase for both phosphotyrosine and phosphothreonine. The haloacid dehalogenase domain at the carboxy terminus contained the tyrosine-phosphatase, and the amino-terminal half carried the threonine-phosphatase. Mutations of the threonine-phosphatase, but not the tyrosine-phosphatase, abolished the ability of EYA4 to enhance the innate immune response, suggesting that EYA regulates the innate immune response by modulating the phosphorylation state of signal transducers for the intracellular pathogens.

We previously showed that *DNase II*^{-/-} mouse embryonic fibroblasts (MEFs) express CXCL10 and IFN- β when they engulf *Cad*^{-/-} apoptotic cells⁴. We used this system to identify molecules involved in the DNA-induced innate immune response. A MEF complementary DNA library (about 20,000 clones) in a retrovirus-based expression vector was prepared, and divided into 400 pools. Retrovirus was produced for each group (see Methods), and used to infect *DNase II*^{-/-} MEFs. The MEFs were incubated with apoptotic *Cad*^{-/-} thymocytes, and CXCL10 levels in the culture supernatant were quantified. One pool of cDNAs that increased the concentration of CXCL10 from 402 to 1,414 pg ml⁻¹ was subjected to sib-selection, which led to the identification of EYA4. When *DNase II*^{-/-} MEFs expressing mouse EYA4 were exposed to apoptotic thymocytes, they produced 3.7- and 25.7-fold more CXCL10 and IFN- β protein, and 8.3- and 11.3-fold more of their transcripts, than the parental cells, respectively (Fig. 1a, b and Supplementary Fig. 1). Overexpressing DNase II in the *DNase II*^{-/-} MEFs blocked the production of CXCL10 and IFN- β , indicating that the effect of EYA4 was on the signal transduction triggered by the undigested DNA.

When wild-type MEFs were transfected with *Escherichia coli* or mammalian DNA, or with polyinosinic-polycytidylc acid (hereafter poly(I:C)), they expressed *Ifnb* messenger RNA^{1,8,9} (Fig. 1c). This response was enhanced 6–20-fold by EYA4. When Namalwa cells were infected with Newcastle disease virus (NDV), they expressed IFN- β in a dose-dependent manner (Fig. 1d): the level of IFN- β was 3.8- and 12.1-fold higher in two *Eya4*-expressing clones than in the parental cells. Accordingly, the titre of Vesicular stomatitis virus (VSV) produced in MEFs expressing EYA4 was about 10 times lower than that in the parental MEF (Fig. 1e). When fetal liver macrophages were intracellularly challenged with poly(I:C), they expressed IFN- β , and this was enhanced by EYA4 (Fig. 1f). On the other hand, the expression of IFN- β by the extracellular poly(I:C), lipopolysaccharide (LPS), or CpG oligonucleotide was not enhanced by EYA4. These results indicated that EYA4 specifically enhanced the innate immune response triggered by intracellular sensors. The mammalian EYA family is comprised of four members¹⁰. The poly(I:C)-induced expression of the *Ifnb* gene was 4–10 times higher in MEFs expressing EYA1, EYA2, EYA3 or EYA4 than in vector-transfected cells (Supplementary Fig. 2). Reduction of EYA expression with a mixture of short-interfering RNAs (shRNAs) against four human EYA genes in 293T cells diminished the poly(I:C)-induced *IFNB* expression, and this could be rescued by the expression of mouse EYA4 that does not carry the target sequence for shRNA (Fig. 1g).

Promoters of the *Ifnb* and *Cxcl10* genes carry interferon regulatory element (IRE) and NF- κ B sites essential for their expression^{11,12}. A deficiency of both *Irf3* and *Irf7* in MEFs completely abrogated the apoptotic-cell-induced production of CXCL10 in EYA4-expressing cells, and co-expression of IRF3 enhanced the effect of EYA4 (Fig. 2a). The poly(I:C)-induced activation of IRF3 (its phosphorylation at Ser 388 (ref. 13) and dimerization¹⁴), as well as the activation of NF- κ B were enhanced by EYA4 (Fig. 2b, c and Supplementary Fig. 3). IPS-1 works as an adaptor^{15,16} for the RIG-I (also known as DDX58) and MDA5 (IFIH1) system; MyD88 and TRIF (TICAM1) are adaptors for the TLR system⁸. The IPS-1-mediated activation of the *Ifnb* promoter¹⁵ was 6.1-fold stronger in EYA4-expressing MEFs than in wild-type MEFs. In contrast, the expression of EYA4 had no effect on the MyD88- and TRIF-induced activation of the *Ifnb* promoter (Fig. 2d). When EYA works as a co-transcription factor for the myogenin (*Myog*) gene, it translocates from the cytoplasm to the nucleus¹⁷. In contrast, EYA4 in MEFs remained in the cytoplasm after stimulation with poly(I:C) (Fig. 2e). When Flag-tagged EYA4, and haemagglutinin (HA)-tagged IPS-1, STING¹⁸ or NLRX1 (refs 19, 20) were expressed in 293T cells, IPS-1, STING and NLRX1 interacted with EYA4 (Fig. 2f and Supplementary Fig. 4). The endogenous IPS-1 interacted with EYA4 transiently after poly(I:C) treatment (Fig. 2g), suggesting that the interaction between EYA4 and IPS-1 is stimulus-dependent.

¹Department of Medical Chemistry, Graduate School of Medicine, Kyoto University, Yoshida-Konoe, Kyoto 606-8501, Japan. ²Solution Oriented Research for Science and Technology, and Core Research for Evolutional Science and Technology, Japan Science and Technology Corporation, Kyoto 606-8501, Japan. [†]Present address: Department of Immunobiology, Yale University School of Medicine, TAC S-669, 300 Cedar Street, New Haven, Connecticut 06510, USA.

*These authors contributed equally to this work.

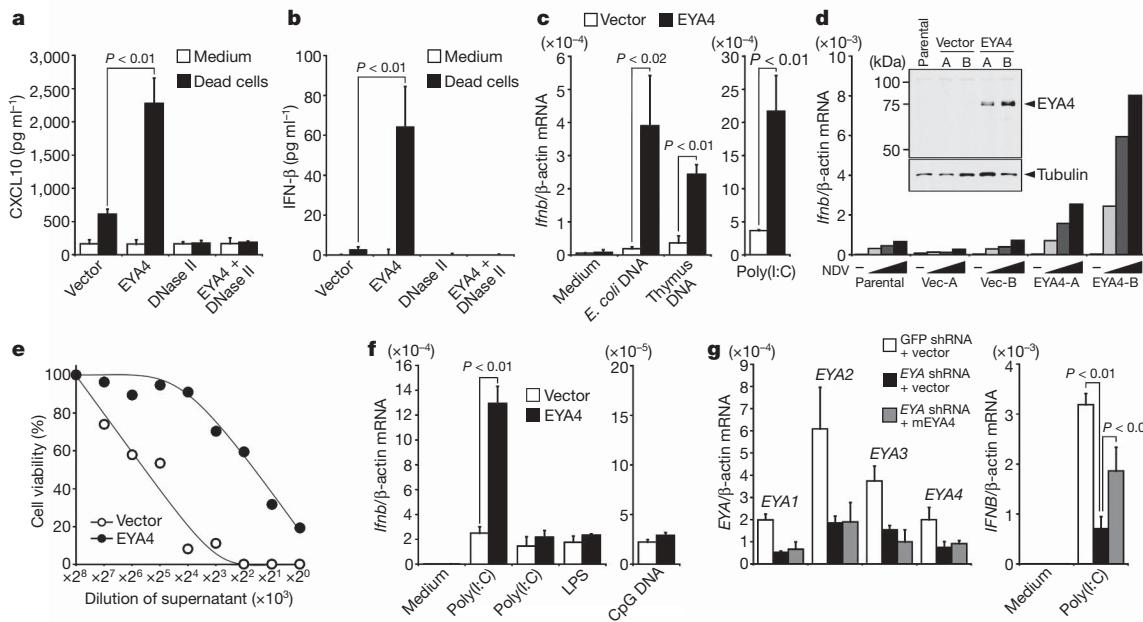


Figure 1 | Identification of EYA4 as a regulator for innate immunity.

a, b, $\text{DNase II}^{-/-}$ MEFs that expressed EYA4, DNase II, both EYA4 and DNase II, or vector control were incubated with medium or dead cells, and the concentrations of CXCL10 (**a**) and IFN- β (**b**) were quantified. **c**, MEFs plus either vector control or EYA4 were transfected with 100 ng ml⁻¹ *E. coli* or calf thymus DNA, or with 10 ng ml⁻¹ poly(I:C), and levels of *Ifnb* mRNA were quantified relative to β -actin mRNA levels. **d**, Parental, and vector- or EYA4-Namalwa cells (two clones each, A and B), were challenged with NDV, and *Ifnb* mRNA was quantified at 8 h. Inset, cell lysates were analysed by western blotting. Tubulin was included as a loading control. **e**, MEFs plus vector

control or EYA4 were infected with 0.001 multiplicity of infection (m.o.i.) of VSV for 12 h. Cytopathic activity in the supernatants was determined. Values are the means from sextuple samples. **f**, Fetal liver macrophages, infected with EYA4 retrovirus, were transfected with 100 ng ml⁻¹ poly(I:C), or extracellularly treated with 100 ng ml⁻¹ of poly(I:C), 1.0 ng ml⁻¹ LPS, or 1.0 μM CpG, and *Ifnb* mRNA levels were quantified at 6 h. **g**, 293T was transfected with shRNA for GFP, or a mixture of shRNA for human EYA1–4, with or without pEF-mEYA4. At 72 h, EYA mRNA was quantified (left). The cells were transfected with 1.0 μg ml⁻¹ of poly(I:C), and *IFNB* mRNA levels were quantified (right). All error bars are s.d.

To characterize the phosphatase activity of EYA, mouse EYA4 was expressed in 293T cells and purified (Fig. 3a). As reported with the C-terminal domain of mouse EYA3 (ref. 6), EYA4 showed tyrosine-phosphatase activity with an artificial peptide under acidic conditions (pH 6.0). Its threonine-phosphatase was potentiated under basic pH, with an optimum pH of around 8.0 (Fig. 3b). The tyrosine-phosphatase required bivalent cations (Mg^{2+} or Mn^{2+}) (Fig. 3c), and ZnCl_2 (1 mM) and EDTA (5 mM) inhibited this activity (Supplementary Fig. 5). On the other hand, the threonine-phosphatase did not require a metal ion (Fig. 3d), and was not inhibited by 10 mM EDTA or EGTA (data not shown), but ZnCl_2 inhibited its activity (Fig. 3d). NaF inhibited both the tyrosine- and threonine-phosphatases, whereas calyculin and okadaic acid inhibited the threonine-phosphatase but not the tyrosine-phosphatase activity (Fig. 3e, f). Under optimal conditions, the K_m and K_{cat} were similar between tyrosine- and threonine-phosphatases, and they were comparable to those for protein phosphatase 2A, protein phosphatase 2B, and protein phosphatase 2C (Supplementary Fig. 6). Other mouse EYA family members, produced in 293T cells (Supplementary Fig. 7), also showed both tyrosine- and threonine-phosphatase activity (Supplementary Fig. 8). None of the members showed phosphatase activity against phosphorylated peptides of (pS)EEEEEE and (pT)EEEEEE, confirming that EYAs are not acid or alkaline phosphatases. The threonine-phosphatase activities of mouse EYA3 and EYA4 were confirmed by preparation in the wheat-germ cell-free system (Supplementary Fig. 9).

The tyrosine-phosphatase activity of the EYA proteins was previously assigned to the haloacid dehalogenase (HAD) domain in the C-terminal half^{6,7}. Although its N-terminal domain is less well conserved, an alignment of the amino acid sequences revealed a conserved motif (Supplementary Fig. 10) in which six tyrosine residues are well conserved. We produced three EYA4 mutants: DYY, Y4 and D352N (Fig. 4a, b). For DYY, Ala was substituted for Asp 246, Tyr 247

and Tyr 250, and for Y4, Ala was substituted for Tyr 258, Tyr 261, Tyr 262 and Tyr 267; for D352N, Asn was substituted for Asp 352. As reported^{6,7}, Asp352Asn lost its tyrosine-phosphatase activity (Fig. 4c), but its threonine-phosphatase activity was unaffected (Fig. 4d). The tyrosine-phosphatase activity of DYY and Y4 was comparable to that of wild-type EYA4, but their threonine-phosphatase activity was severely affected, with Y4 showing the most disruption. This was also true with peptides from the C-terminal region of RNA polymerase II (Supplementary Fig. 11). When EYA3 was divided into an N-terminal and C-terminal domain, the C-terminal domain carried the tyrosine-phosphatase, but not the threonine-phosphatase, activity (Supplementary Fig. 12). In contrast, its N-terminal domain showed threonine-phosphatase activity, but negligible tyrosine-phosphatase activity. These results apparently contradict the previous report that assigned the threonine-phosphatase of EYA to the HAD domain⁵. When Asp352Asn and Y4 mutants were expressed in mouse fetal liver macrophages (Fig. 4e), Asp352Asn enhanced poly(I:C)-induced IFN- β expression as efficiently as the wild-type EYA4, whereas Y4 severely lost this ability (Fig. 4f). This was in contrast to the effect of the Y4 mutation on the ability of EYA4 to enhance the SIX4-mediated transcription of the myogenin gene promoter (Supplementary Fig. 13). The point mutations in the N or C termini did not affect its ability to interact with IPS-1 (Fig. 4g), suggesting that the threonine-phosphatase is critical for the ability of EYA4 to enhance the innate immune response.

Here we have reported that EYA4 interacts with IPS-1, STING and NLRX1, which are on the mitochondrial outer membrane or endoplasmic reticulum^{15,18,19,21}, and stimulates the IRF3-mediated transcription of the *Ifnb* and *Cxcl10* genes. A recent report indicates that EYA dephosphorylates a tyrosine residue of histone H2AX, and renders the cells resistant to genotoxic-agent-induced apoptosis²². These results indicate that EYA has dual functions: one to regulate the chromatin structure using its tyrosine-phosphatase, the other to

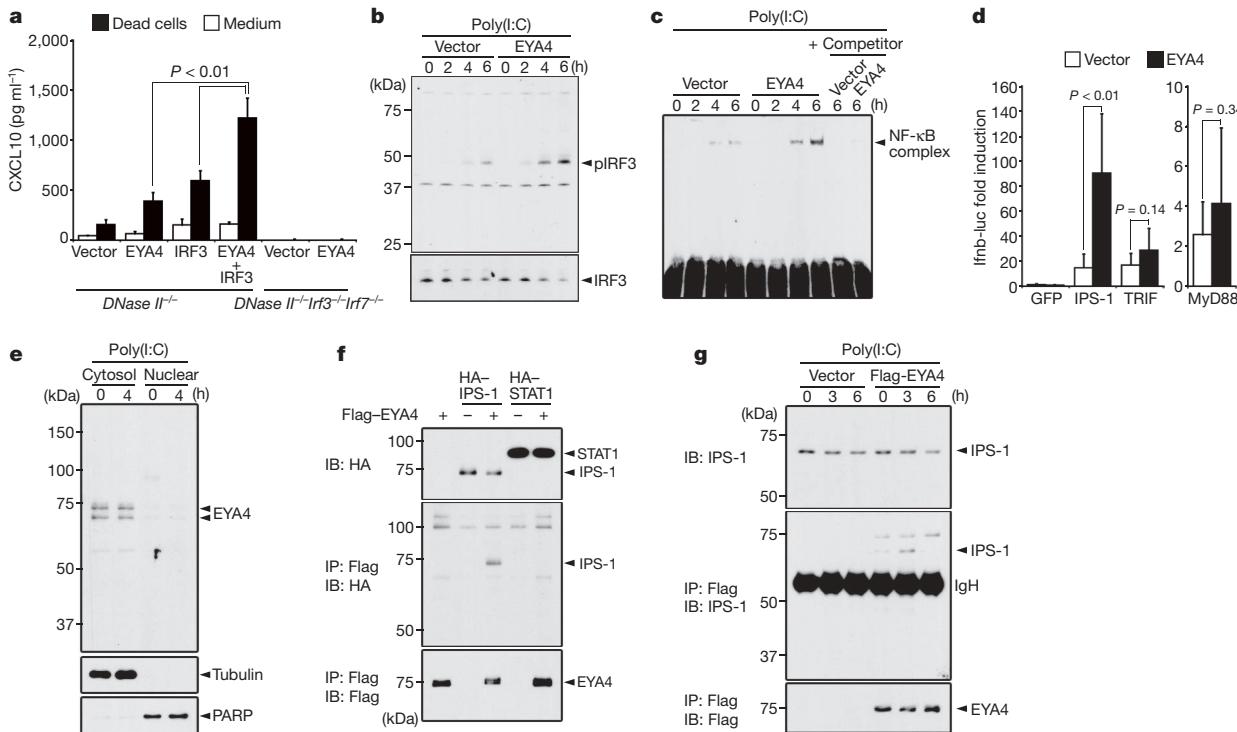


Figure 2 | Activation of the signalling cascade by EYA4. **a**, *DNase II^{-/-}* or *DNase II^{-/-} Irf3^{-/-} Irf7^{-/-}* MEFs expressing EYA4 and/or IRF3, were cultured with medium or dead cells, and CXCL10 levels were determined. **b, c**, MEFs containing vector control or EYA4 were transfected with 1 $\mu\text{g ml}^{-1}$ poly(I:C) for the indicated times. The cell lysates were analysed by western blot for phosphorylated-IRF3 (pIRF3) (**b**), or nuclear extracts were analysed by electrophoretic mobility shift assays (EMSA) for NF-κB (**c**). **d**, MEFs containing vector control or EYA4 were transfected with pGL3-Ifnb-luc, and pEF-EGFP, pEF-IPS-1, pEF-TRIF or pEF-MyD88. The luciferase activity is shown relative to the EGFP value. **e**, MEFs were treated with 1.0 $\mu\text{g ml}^{-1}$ poly(I:C) for 4 h, and cell extracts were fractionated into

cytoplasmic and nuclear fractions and analysed by western blotting. Two bands for EYA4 are probably due to alternative splicing²⁹. The fractionation of the cell extracts was confirmed by blotting for tubulin and PARP. **f**, 293T cells were co-transfected with the vector for Flag-EYA4 and HA-tagged IPS-1 and STAT1. Cell lysates were immunoprecipitated (IP) with anti-Flag and analysed by western blotting. Cell lysates are shown at the top (5% of the input). IB, immunoblot. **g**, MEFs plus vector control or EYA4 were transfected with 1.0 $\mu\text{g ml}^{-1}$ poly(I:C) for the indicated times. Cell lysates were immunoprecipitated with anti-Flag and analysed by western blotting. Cell lysates are shown at the top (10% of input). All error bars are s.d.

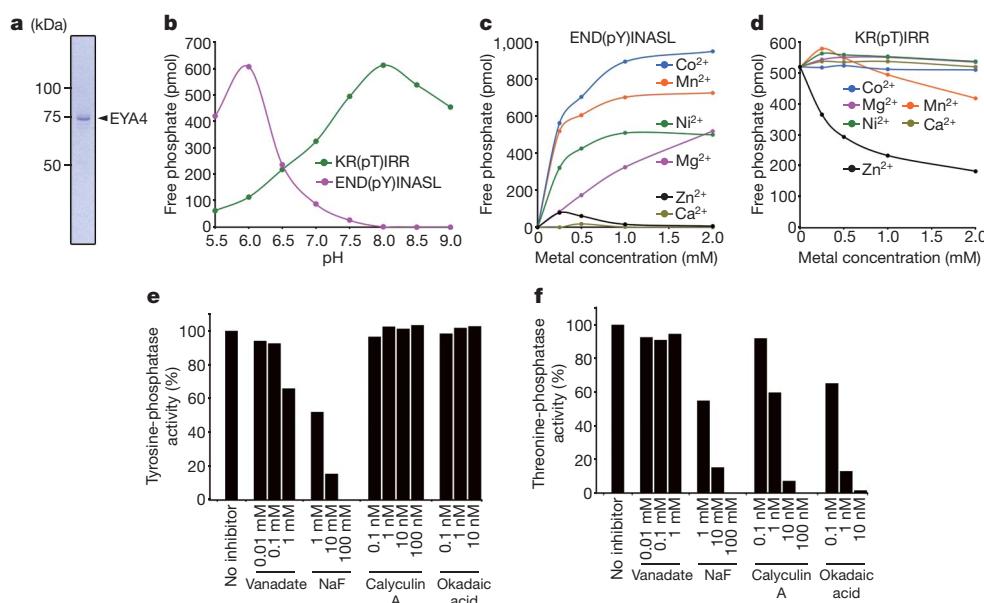


Figure 3 | Two different phosphatase activities in mouse EYA. **a**, Mouse EYA4 expressed in 293T, was purified and analysed by SDS-PAGE. **b-d**, Dephosphorylation with END(pY)INASL and 5 pmol EYA4 (**b, c**), or with KR(pT)IRR and 2.5 pmol EYA4 (**b, d**) at the indicated pH (**b**) or in the

presence of metal ions (**c, d**) is shown. **e, f**, Dephosphorylation of RRLIEDAE(pY)AARG (**e**) or SDQEKRKQI(pT)VRGL (**f**) by 1.25 pmol EYA4 in the presence of phosphatase inhibitors is shown.

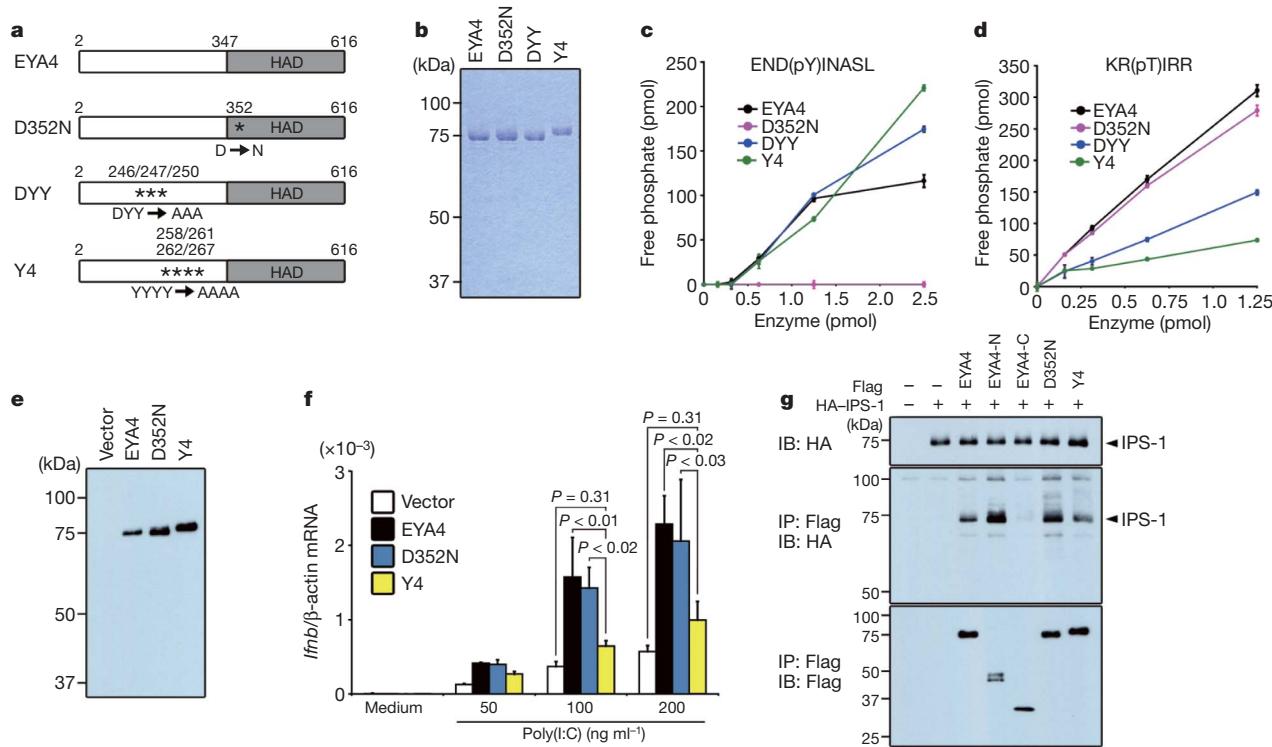


Figure 4 | The threonine-phosphatase activity of EYA4 is required for the innate immune reaction. **a**, Diagrams of EYA4 and its mutants are shown. **b**, EYA4 and its mutants, produced in 293T cells, were purified and analysed by SDS-PAGE. **c, d**, Dephosphorylation by EYA4 and its mutants is shown. **e, f**, Fetal liver macrophages were infected with EYA4 retrovirus. Cell lysates were analysed by western blot using an anti-Flag antibody (M2) (**e**). The cells

were transfected with poly(I:C) for 6 h, and *Ifnb* mRNA levels were quantified relative to β-actin levels (**f**). **g**, 293T cells transfected with vectors for Flag-EYA4 or its mutants, and with HA-IPS-1. Cell lysates were immunoprecipitated (IP) with anti-Flag, followed by western blotting. Shown at the top is 10% of the input. All error bars are s.d.

regulate the innate immune response in the cytoplasm using its threonine-phosphatase. It will be interesting to study how the two phosphatase activities of EYA are regulated, and how the cellular localization of EYA is determined. The N-terminal region of EYA for the threonine-phosphatase had no apparent similarity with other phosphatases, yet its activity could be inhibited by okadaic acid. As found between protein phosphatase 1 and protein phosphatase 2A²³, the tertiary structure of the N-terminal domain of EYA may be similar to other threonine phosphatases. Several kinases (RIP1 (also known as RIPK1), IKK-ε (IKBKE), TBK1, IKK-α (CHUK) and IKK-β (IKBKB)) are involved in RIG-I-MDA5-mediated signal transduction²⁴. It was previously proposed²⁵ that binding of viral RNA to RIG-I-MDA5 recruits signalling molecules (TRADD, TRAF3 and NEMO (IKBKG)), to form a large complex. It is possible that EYA is one of the components of this large complex, and regulates the signalling by modulating their phosphorylation state (Supplementary Fig. 14). Genes in the EYA family are often mutated in various autosomal-dominant disorders associated with branchial arch malformations, hearing loss, and ocular segment anomalies²⁶. Mutations can be found throughout the EYA molecule²⁷, and some of these mutations have no effect on the tyrosine-phosphatase activity or on the interaction with a SIX family member²⁸. Whether any of the mutations found in genes of the EYA family in human patients affect the threonine-phosphatase activity remains to be determined.

METHODS SUMMARY

Expression cloning. Double-stranded cDNA with poly(A) RNA from MEFs was inserted into a retroviral vector, and introduced into *E. coli* to generate a library (400 pools of 50 clones). Plasmid DNA from each pool was introduced into packaging cells, and the culture supernatant was used to infect *DNase II*^{-/-} MEFs. The MEFs were allowed to engulf apoptotic *Cad*^{-/-} thymocytes, and CXCL10 in the culture supernatant was quantified by ELISA. The positive pools were subjected to sib-selection.

Molecular biology, biochemistry and cell biology. Macrophages and MEFs were transformed by retrovirus-mediated transfection. Namalwa cells expressing mouse EYA4 were established by electroporation with EYA4 expression vector. The MFG-E8-mediated system was used for engulfment of apoptotic cells. The Fugene system was used to introduce nucleic acids into cells. IFN-β and CXCL10 were quantified by ELISA. The promoter activity of the *Ifnb* and myogenin genes was determined with Dual-Luciferase Reporter Assay System (Promega). Real-time PCR was performed using a LightCycler 480 system (Roche Diagnostics). The shRNA expression plasmids for human *EYA1*, *EYA2*, *EYA3* and *EYA4* were purchased from OriGene.

Production of recombinant EYAs. For recombinant EYA, the Flag-tag was ligated at the 5' end of *Eya*, and inserted into the pEF-BOS vector. A point mutation was introduced by recombinant PCR. The expression vector was introduced into human 293T cells, and the recombinant protein was purified using anti-Flag M2 gel. In some cases, EYA was synthesized using the wheat-germ cell-free system, and purified.

Phosphatase assay. Phosphorylated synthetic peptides were custom-synthesized at the Toray Research Center, MBL, and Invitrogen. To assay the phosphatase activity, EYA was incubated with 400 μM phosphorylated peptide at 37 °C for 60 min, and released phosphate was quantified by a colourimetric method using malachite green-molybdate.

Statistical analysis. Results were statistically analysed using an analysis of variance (ANOVA) test.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 25 March; accepted 8 May 2009.

Published online 28 June; corrected 23 July 2009 (see full-text HTML version for details).

1. Meylan, E., Tschopp, J. & Karin, M. Intracellular pattern recognition receptors in the host response. *Nature* 442, 39–44 (2006).
2. Yoshida, H., Okabe, Y., Kawane, K., Fukuyama, H. & Nagata, S. Lethal anaemia caused by interferon-β produced in mouse embryos carrying undigested DNA. *Nature Immunol.* 6, 49–56 (2005).
3. Kawane, K. et al. Chronic polyarthritis caused by mammalian DNA that escapes from degradation in macrophages. *Nature* 443, 998–1002 (2006).

4. Okabe, Y., Kawane, K., Akira, S., Taniguchi, T. & Nagata, S. Toll-like receptor-independent gene induction program activated by mammalian DNA escaped from apoptotic DNA degradation. *J. Exp. Med.* **202**, 1333–1339 (2005).
5. Li, X. et al. Eya protein phosphatase activity regulates Six1-Dach-Eya transcriptional effects in mammalian organogenesis. *Nature* **426**, 247–254 (2003).
6. Rayapureddi, J. P. et al. Eyes absent represents a class of protein tyrosine phosphatases. *Nature* **426**, 295–298 (2003).
7. Tootle, T. L. et al. The transcription factor Eyes absent is a protein tyrosine phosphatase. *Nature* **426**, 299–302 (2003).
8. Akira, S., Uematsu, S. & Takeuchi, O. Pathogen recognition and innate immunity. *Cell* **124**, 783–801 (2006).
9. Medzhitov, R. Recognition of microorganisms and activation of the immune response. *Nature* **449**, 819–826 (2007).
10. Jemc, J. & Rebay, I. The eyes absent family of phosphotyrosine phosphatases: properties and roles in developmental regulation of transcription. *Annu. Rev. Biochem.* **76**, 513–538 (2007).
11. Kim, T. K. & Maniatis, T. The mechanism of transcriptional synergy of an *in vitro* assembled interferon- β enhanceosome. *Mol. Cell* **1**, 119–129 (1997).
12. Ohmori, Y. & Hamilton, T. A. Cooperative interaction between interferon (IFN) stimulus response element and kB sequence motifs controls IFN γ - and lipopolysaccharide-stimulated transcription from the murine IP-10 promoter. *J. Biol. Chem.* **268**, 6677–6688 (1993).
13. Sharma, S. et al. Triggering the interferon antiviral response through an IKK-related pathway. *Science* **300**, 1148–1151 (2003).
14. Iwamura, T. et al. Induction of IRF-3/-7 kinase and NF- κ B in response to double-stranded RNA and virus infection: common and unique pathways. *Genes Cells* **6**, 375–388 (2001).
15. Kawai, T. et al. IPS-1, an adaptor triggering RIG-I- and Mda5-mediated type I interferon induction. *Nature Immunol.* **6**, 981–988 (2005).
16. Meylan, E. et al. Cardif is an adaptor protein in the RIG-I antiviral pathway and is targeted by hepatitis C virus. *Nature* **437**, 1167–1172 (2005).
17. Ohto, H. et al. Cooperation of six and eya in activation of their target genes through nuclear translocation of Eya. *Mol. Cell. Biol.* **19**, 6815–6824 (1999).
18. Ishikawa, H. & Barber, G. N. STING is an endoplasmic reticulum adaptor that facilitates innate immune signalling. *Nature* **455**, 674–678 (2008).
19. Moore, C. B. et al. NLRX1 is a regulator of mitochondrial antiviral immunity. *Nature* **451**, 573–577 (2008).
20. Tattoli, I. et al. NLRX1 is a mitochondrial NOD-like receptor that amplifies NF- κ B and JNK pathways by inducing reactive oxygen species production. *EMBO Rep.* **9**, 293–300 (2008).
21. Jin, L. et al. MPYS, a novel membrane tetraspanner, is associated with major histocompatibility complex class II and mediates transduction of apoptotic signals. *Mol. Cell. Biol.* **28**, 5014–5026 (2008).
22. Cook, P. J. et al. Tyrosine dephosphorylation of H2AX modulates apoptosis and survival decisions. *Nature* **458**, 591–596 (2009).
23. Das, A. K., Helps, N. R., Cohen, P. T. & Barford, D. Crystal structure of the protein serine/threonine phosphatase 2C at 2.0 Å resolution. *EMBO J.* **15**, 6798–6809 (1996).
24. Lee, M. S. & Kim, Y. J. Signaling pathways downstream of pattern-recognition receptors and their cross talk. *Annu. Rev. Biochem.* **76**, 447–480 (2007).
25. Michallet, M. C. et al. TRADD protein is an essential component of the RIG-like helicase antiviral pathway. *Immunity* **28**, 651–661 (2008).
26. Abdelhak, S. et al. A human homologue of the *Drosophila* eyes absent gene underlies branchio-oto-renal (BOR) syndrome and identifies a novel gene family. *Nature Genet.* **15**, 157–164 (1997).
27. Orten, D. J. et al. Branchio-oto-renal syndrome (BOR): novel mutations in the EYA1 gene, and a review of the mutational genetics of BOR. *Hum. Mutat.* **29**, 537–544 (2008).
28. Mutsuddi, M. et al. Using *Drosophila* to decipher how mutations associated with human branchio-oto-renal syndrome and optical defects compromise the protein tyrosine phosphatase and transcriptional functions of eyes absent. *Genetics* **170**, 687–695 (2005).
29. Borsani, G. et al. EYA4, a novel vertebrate gene related to *Drosophila* eyes absent. *Hum. Mol. Genet.* **8**, 11–23 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank T. Fujita and M. Yoneyama for the NDV, A. Sehara-Fujisawa for the myogenin promoter, and M. Fujii and M. Harayama for secretarial assistance. This work was supported in part by Grants-in-Aid from the Ministry of Education, Science, Sports, and Culture in Japan. T.S. is a Research Assistant for Kyoto University Global COE program (Center for Frontier Medicine).

Author Contributions Y.O. screened the cDNA library, and identified EYA4 as a regulator of the innate immune reaction. T.S. produced recombinant EYA in 293T cells and wheat-germ extracts, biochemically characterized, and found its threonine-phosphatase activity. S.N. designed the research and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to S.N. (snagata@mfour.med.kyoto-u.ac.jp).

METHODS

Cells, virus and other reagents. Monkey Vero cells were cultured in αMEM containing 15% FBS. MEFs and human 293T cells were cultured in DMEM containing 10% FBS. Human Namalwa cells were maintained in RPMI1640 containing 10% FBS. *DNase II*^{-/-} MEFs expressing α_vβ₃-integrin (*DNase II*^{-/-}MEFs/integrin) were described previously⁴. *DNase II*^{-/-} *Irf3*^{-/-} *Irf7*^{-/-} MEFs were established from embryonic day (E)13.5 *DNase II*^{-/-} *Irf3*^{-/-} *Irf7*^{-/-} mouse embryos as described⁴. To prepare the fetal liver macrophages, the liver from E14.5 mouse embryo was dissected and passed through a 22-gauge needle five times. After washing with PBS, the cells were cultured in αMEM containing 10% FBS supplemented with 10% (v/v) supernatant of CMG14-12 cells producing mouse macrophage colony-stimulating factor³⁰. After 48 h, the cells were washed, and adherent cells were used as fetal liver macrophages. NDV was provided by T. Fujita (Kyoto University) and propagated in fertilized chicken eggs as described³¹. VSV (Indiana strain) was propagated in mouse L929 cells.

Antibodies used were mouse monoclonal antibodies against Flag (M2 antibody, Sigma), HA (clone 12C5, Boehringer Mannheim; clone 16B12, Babco), PARP (BD), and anti-tubulin (EMD Chemicals), rabbit polyclonal antibodies against IRF3 (Zymed Laboratories) and IPS-1 (Cell Signaling), and rabbit monoclonal antibody against phosphorylated-IRF3 (Ser 396; Cell Signaling).

Leucine zipper-tagged human Fas ligand³² and mouse MFG-E8 (ref. 33) were produced in monkey COS7 and human 293T cells, respectively. Poly(I:C) was purchased from GE Healthcare. Calf thymus DNA, *E. coli* DNA and LPS were from Sigma. A phosphorothioate-stabilized CpG oligonucleotide³⁴ of 5'-TCCATGACGTTCTGATGC-3' was purchased from Hokkaido System Science.

Production of monoclonal antibody against mouse EYA4. To produce monoclonal antibody against mouse EYA4, the N-terminal domain of EYA4 (amino acids 2–346) was joined to maltose-binding protein (MBP) using pMAL-p2X vector (New England BioLab), or to glutathione-S-transferase (GST) using pGEX-5X-1 vector (GE Healthcare). MBP-EYA4 and GST-EYA4 were produced in *E. coli* BL21(DE3) pLysS strain, and purified using Amylose-resin (New England BioLab) and glutathione-sepharose (GE Healthcare), respectively. Armenian hamsters were immunized by subcutaneously injecting four times with 50 µg of MBP-EYA4 in a 2-week interval. For the first injection, MBP-EYA4 was mixed with complete Freund's adjuvant (DIFCO), while it was mixed with incomplete Freund's adjuvant for the second to fourth injections. A final injection with 50 µg MBP-EYA4 was given into the foot pads. Two days later, popliteal and inguinal lymph nodes were collected, and the lymphocytes were fused with mouse myeloma NSO cell line³⁵. Hybridomas were selected in HAT medium (Boehringer Ingelheim), and the supernatants were tested by ELISA for their ability to recognize GST-EYA4. Positive clones were examined for their suitability in western blotting, and one hybridoma (clone 5-7C 48) was subjected to limiting dilution. The hybridoma was grown in GIT medium (Nihon Pharmaceutical), and the antibodies were purified by protein A-sepharose.

Expression cloning. Poly(A) RNA was prepared from MEFs using an mRNA purification kit (GE Healthcare). Double-stranded cDNA was generated with the SuperScript Choice System (Invitrogen). cDNAs longer than 0.5 kb were ligated into a BstXI-digested pMXs vector³⁶ using the BstXI adaptor, and introduced into *E. coli* DH10B by electroporation, to generate a cDNA library of about 20,000 clones. The library was divided into 400 pools of 50 clones, and plasmid DNA was prepared from each pool using NucleoSpin Multi-96 Plus Plasmid (Macherey-Nagel). The DNA was introduced into Plat-E packaging cells³⁷, and the culture supernatant was used to infect *DNase II*^{-/-} MEFs/integrin. At 48 h after the infection, the MEFs were allowed to engulf apoptotic *Cad*^{-/-} thymocytes, and the concentration of CXCL10 in the culture supernatant was determined by ELISA. The pools that gave a positive result were subjected to sub-selection.

Engulfment of apoptotic cells and transfection of DNA. The MFG-E8-dependent engulfment of apoptotic cells was performed as described³³. In brief, thymocytes from *Cad*^{-/-} mice were treated for 2 h with Fas ligand to induce apoptosis in about 50% of the cell population. The apoptotic cells were added to MEFs at a ratio of 50:1, and incubated for 20 h in the presence of 0.1 µg ml⁻¹ MFG-E8.

To introduce nucleic acids into cells, the nucleic acids were preincubated at room temperature for 15 min with FuGENE6 transfection reagent (Roche Diagnostics) at a ratio of 1:1.5 (w/v) in DMEM, added to the cells, and incubated at 37 °C for 6 (for macrophages) or 12 (for MEF and 293T) h.

Enzyme-linked immunosorbent assay. The concentrations of IFN-β and CXCL10 were determined by ELISAs using the mouse Interferon Beta ELISA Kit (PBL) and mouse CXCL10 DuoSet ELISA Development system (R&D), respectively. Streptavidin-conjugated horseradish peroxidase (HRP) or alkaline phosphatase was used as the substrate. In some cases, signals were amplified with the AmpliQ (Dako) detection system.

Real-time PCR. Total RNA was prepared from cells using the RNeasy kit with the RNase-Free DNase set (Qiagen). After reverse-transcription with oligo(dT) primer using a Transcripter First Strand cDNA Synthesis Kit (Roche Diagnostics), aliquots of the products were amplified using the LightCycler 480 system (Roche Diagnostics) according to the instructions provided by the manufacturer. The primers used for the mouse genes were as follows: *Ifnb*, 5'-CCACCACAGCCCTCTCATCAACTAT-3' and 5'-CAAGTGGAGAGCAG TTGAGGACATC-3'; *Cxdl10*, 5'-CCATCAGCACCATGAACCCAAGT-3' and 5'-CACTCCAGTTAAGGAGCCCTTTAGACC-3'; β-actin, 5'-TGTGATGGT GGGATGGGTCAAGT-3', and 5'-TTTGATGTCACGCACGATTCC-3'. The primers for the human genes were as follows: *IFNB*, 5'-CCAACAA GTGTCTCCTCCAAA-3' and 5'-CCTCAGGGATGTCAAAGTCA-3'; *EYA1*, 5'-AGGCCACCATACAGCTACCAGAT-3' and 5'-GCTGGTCATATAATGTC TGGA-3'; *EYA2*, 5'-GGCACTAACCTCATCAACTCC-3' and 5'-CACCACG TAGACAGCTTTCTG-3'; *EYA3*, 5'-CCAGATGTCAGTGATCAGAACG-3' and 5'-GGGTAGACAGCATAGGGTTGAG-3'; *EYA4*, 5'-CGCTTGTGTACGC TATTGTTGT-3' and 5'-TTTCTCACTTCTCCTGCCACTC-3'; β-actin, 5'-GCATCCTCACCTGAAGTAC-3' and 5'-CTTAATGTCACGCACGATT TC-3'.

Expression plasmid. The cDNAs of mouse EYA1 (GenBank accession NP_034294), EYA2 (GenBank accession NP_034295), EYA3 (GenBank accession NP_034296), and EYA4 (GenBank accession NP_034297) were isolated by RT-PCR from the RNA of MEFs, and their authenticity was confirmed by DNA sequencing. As the N- and C-terminal regions of EYA3, amino acids 2–255 (encoded by exons 2–10), and amino acids 256–526 (encoded by exons 11–18), were used, respectively. Whereas amino acids 2–345 (encoded by exons 2–11), and amino acids 346–616 (encoded by exons 12–19), were used as the N- and C-terminal regions of EYA4, respectively. In the EYA4-Y4 mutant, tyrosine was replaced by alanine at four positions (amino acids 258, 261, 262 and 267). The EYA4-DYY mutant was produced by replacing Asn at position 246 and two Tyr residues at positions 247 and 250 with Ala. In the Asp352Asn mutant, the Asp at position 352 was replaced with Asn. The expression plasmid for these mutants were constructed by recombinant PCR³⁸. All EYA proteins and their derivatives were Flag-tagged at the N terminus.

Mouse cDNAs for *IPS-1* (GenBank accession NP_659137), *Trif* (GenBank accession NP_778154), *Sting* (GenBank accession XP_905219), *Nlrx1* (GenBank accession NP_848507), *Irf3* (GenBank accession NP_058545), *Six4* (GenBank: NP_035512) and *Myd88* (GenBank: NP_034981) were isolated by RT-PCR from the RNA of MEF or fetal liver macrophages. Each cDNA was tagged with three tandem haemagglutinins (a peptide of influenza haemagglutinin, YPYDVPDYA) at the N terminus (IPS-1, TRIF, SIX4, NLRX1 and Myd88) or the C terminus (STING), or with Flag at the N terminus (IRF3).

Production of recombinant EYA proteins. The mouse EYA proteins were produced by transfecting human 293T cells with the respective expression plasmids, and the accumulated protein was purified using anti-Flag M2 affinity gel (Sigma). In brief, the expression vectors for Flag-tagged EYA and its derivatives were generated in pEF-BOS vector³⁹ and introduced into human 293T cells using the calcium-phosphate co-precipitation method. Sixty to seventy-two hours later, the cells were lysed by incubating them for 30 min at 4 °C in 50 mM Tris-HCl (pH 7.4) containing 1.0% (v/v) NP40, 150 mM NaCl, 2 mM EDTA, 2 mM EGTA, 2.0 µg ml⁻¹ pepstatin, 2.0 µg ml⁻¹ leupeptin, 2 µg ml⁻¹ aprotinin, and 2 µM pABSF (4-(2-aminoethyl)-benzenesulfonyl fluoride hydrochloride). After removing the cell debris by centrifugation at 39,800g for 40 min, the cell extracts were affinity-purified with anti-Flag-Sepharose (M2, Sigma), and the protein was eluted with 50 mM Tris-HCl buffer (pH 7.4) containing 100 µg ml⁻¹ of Flag peptide (Sigma), 150 mM NaCl, and 0.1% NP40.

Recombinant EYA3 and EYA4 were also produced *in vitro* using the wheat-germ cell-free system⁴⁰. In brief, the DNA fragment for the Flag-tagged EYA3 or EYA4 was inserted into pEU E01 vector (Cell Free Science). With the resultant plasmid as a template, *Eya3* or *Eya4* mRNA was synthesized with SP6 RNA polymerase using a kit from Cell Free Sciences according to the instructions provided by the manufacturer. EYA protein was then synthesized with the mRNA by incubation at 15 °C for 20 h with 250 µl of wheat-germ extracts and 5.5 ml of SUM-AMIX (Cell Free Sciences). After incubation, the reaction mixture was diluted eightfold with 50 mM Tris-HCl (pH 7.4) containing 1.0% NP40, 150 mM NaCl, 2 mM EDTA, 2 mM EGTA, 2.0 µg ml⁻¹ pepstatin, 2.0 µg ml⁻¹ leupeptin, 2 µg ml⁻¹ aprotinin and 2 µM pABSF, and the Flag-tagged EYA protein was purified using anti-Flag-Sepharose as described earlier.

Assay for phosphatase activity. The phosphorylated synthetic peptides KR(pS)IRR, KR(pT)IRR, RRA(pT)VA and END(pY)INASL were custom-synthesized by the Toray Research Center, and KR(pT)IRR, DLDVPIPGRF DRRV(pT)VAAE, SDQEKRKQI(pT)VRGL, LDPRQVEMIRRRP(pT)PAML, (pT)EEEEEE, (pS)EEEEEE, DADE(pY)LIPQQG, END(pY)INASL, RRLIEDAE (pY)AARG, TRDI(pY)ETDYRK and DRV(pY)IHPF were custom-synthesized

by MBL. YSPTSPS(pY)SPTSPS, YSPTSPSY(pS)PT(pS)PS and YSPTSPSY SP(pT)SPS were custom-synthesized by Invitrogen. Sodium orthovanadate, okadaic acid and calyculin A were purchased from EMD Chemicals. Recombinant human protein phosphatase 2A, 2B and 2C were purchased from Wako, Promega and EMD Chemicals, respectively.

To assay the phosphatase activity, the recombinant EYA was incubated in a final volume of 50 µl with 400 µM phosphorylated peptide at 37 °C for 60 min in 50 mM MES-KOH buffer (pH 6.0) containing 2 mM MgCl₂ and 50 µM dithiothreitol (DTT) for the tyrosine phosphatase or in 50 mM Tricine-KOH buffer (pH 8.0) containing 5 mM EDTA and 50 µM DTT for the serine/threonine-phosphatase. The phosphatase activity of human protein phosphatase 2A, 2B and 2C was assayed similarly by incubating at 37 °C for 60 min. The buffer was essentially as previously reported^{41,42}. That is, 50 mM Tris-HCl buffer (pH 7.0) containing 0.1 mM EGTA and 14.2 mM 2-mercaptoethanol for protein phosphatase 2A, 50 mM HEPES-KOH buffer (pH 7.5) containing 2 mM NiCl₂ for protein phosphatase 2B, and 50 mM Tris-HCl buffer (pH 7.5) containing 0.1 mM EDTA, 30 mM MgCl₂ and 1 mM DTT for protein phosphatase 2C. The released phosphate was quantified by a colourimetric method using the malachite green-molybdate as described⁴³. In brief, the reaction was terminated by the addition of 50 µl of H₂O and 20 µl of 1.75% (w/v) ammonium molybdate·4H₂O in 6.0 N H₂SO₄, and incubated at room temperature for 20 min. A solution (20 µl) containing 0.35% (w/v) malachite green and 0.35% (w/v) polyvinylalcohol (molecular mass 13,000–23000 Da) was added to the reaction mixture, which was further incubated for 20 min at room temperature. The malachite green-ammonium molybdate phosphate complex was detected at 620 nm using a Micro Plate Reader (BioLumin 960). *K_m* and *K_{cat}* values for the phosphatase activities of EYA were determined using Hanes-Woolf Plot.

Stable cell transformants. To express EYA and IRF3 in mouse MEFs and fetal liver macrophages, the corresponding cDNA was introduced into pMXs vector³⁶. Retrovirus was prepared by transfecting PlatE packaging cells³⁷ with the expression vector, and used to infect MEF and macrophages⁴. To establish Namalwa cells that constitutively express mouse EYA4, the Flag-tagged EYA4 cDNA was introduced into pEF-BOS vector³⁹ to generate pEF-Flag-EYA4. Namalwa cells were co-transfected by electroporation with Aat II-digested pEF-Flag-EYA4 and a plasmid carrying the hygromycin-resistant gene. After being selected with 1.0 mg ml⁻¹ hygromycin, the clones expressing EYA4 were identified by western blotting with anti-Flag antibody.

Cytopathic assay for VSV. Samples carrying VSV were seriously diluted, and incubated at 37 °C for 40 h with 1 × 10⁴ Vero cells in 96-well microtitre plate. After incubation, surviving cells were stained with 0.5% crystal violet, and dissolved in methanol. Absorbance at 595 nm was determined using an ELISA reader (Infinite 2000, TECAN).

shRNA. shRNA expression plasmids for human *EYA1*, *EYA2*, *EYA3* and *EYA4* in a pRS shRNA vector carrying a puromycin-resistance gene were purchased from OriGene. The target sequences for the shRNAs were as follows: for human *EYA1*, 5'-ATCTGGCATCACCAGCCAAGCAGTACAG-3'; for human *EYA2*, 5'-CC GAGTCACTTGCTGGTGAATAACACACA-3'; for human *EYA3*, 5'-AGT AGGCCAGCATCTAAACCAAGGATTATC-3'; for human *EYA4*, 5'-CCTG AGGTTTGAATCTGACTTACGGCGTGCAGTGC-3'; and for the control GFP, 5'-TGAC CACCCCTGACCTACGGCGTGCAGTGC-3'. The U6-EYA1 and U6-EYA2 portions were joined into one pRS vector, and the U6-Eya3 and U6-EYA4 portions were joined into another pRS vector. Human 293T cells were transfected using the Fugene transfection reagent with the two shRNA plasmids. Sixty hours later, the cells were transfected with poly(I:C), and the *Ifnb* mRNA was quantified.

Luciferase assay. The promoter activity was determined by luciferase assay using the Dual-Luciferase Reporter Assay System (Promega). In brief, a DNA fragment carrying the mouse *Ifnb* promoter region (-121 to +22; +1, transcription start site)⁴⁴ was amplified by PCR from mouse genomic DNA. The mouse myogenin promoter (-182 to +52)⁴⁵ was provided by A. Sehara-Fujisawa (Kyoto University). The *Ifnb* and myogenin promoter was inserted upstream of the firefly luciferase gene of pGL3 to generate pGL3-mIFNβ125luc and pGL3MG-185, respectively. The expression vector for EYA4, IPS-1, TRIF, MyD88 or Six4 cDNA in pEF-BOS was introduced into MEF or 293T cells together with pGL3-mIFNβ125luc or pGL3MG-185 reporter plasmid. Twenty-four (for *Ifnb* promoter) or forty-eight (for myogenin promoter) hours later, the luciferase activity was determined as described previously⁴⁶. pEF-Rluc or pSV-Rluc, which carries the *Renilla* luciferase gene downstream of the elongation factor 1 promoter⁴⁶ or SV40-early promoter, was used as an internal control for transfection.

Electrophoretic mobility shift assay. For the EMSAs, nuclear extracts were prepared as described⁴⁷. The probe DNA carrying the NF-κB binding site (5'-AGTGAGGGACTTCCCAGG-3') was labelled with biotin using a biotin 3' end DNA labelling kit (Pierce). For the binding reaction, nuclear extracts (2 µg protein) were incubated for 30 min at 25 °C with 20 fmol of biotin-labelled probe in 10 µl of a reaction mixture containing 10 mM Tris-HCl (pH 7.5), 50 mM KCl,

1 mM DTT, 5% glycerol, 2% polyvinylalcohol, and 500 ng poly(I:C) in the presence or absence of 4 pmoles of unlabelled probe. The reaction mixture was separated by electrophoresis on a 6% polyacrylamide gel in 50 mM Tris-HCl (pH 8.5) containing 380 mM glycine, and 2 mM EDTA, and the shifted band was detected with the LightShift Chemiluminescent EMSA Kit (Pierce).

Fractionation of the cell extracts. The cell extracts were fractionated into nuclear and cytoplasmic fractions as described previously⁴⁷. In brief, MEFs were incubated at 4 °C for 10 min in hypotonic buffer (10 mM HEPES-KOH buffer (pH 7.9), 0.1 mM EDTA, 0.1 mM EGTA, 1 mM DTT, 10 mM KCl and a cocktail of protease inhibitors and phosphatase inhibitors (Roche Diagnostics) to allow cells to swell. The cells were then lysed by vigorously vortexing in the hypotonic buffer containing 6% sucrose and 0.1% NP-40. The lysates were spun at 14,000g for 2 min, and the supernatants were used as the cytoplasmic fraction. The pellets were suspended in 20 mM HEPES-KOH (pH 7.9) containing 0.2 mM EDTA, 2 mM EGTA, 2 mM DTT, 25% glycerol, 420 mM NaCl, and a cocktail of protease inhibitors and phosphatase inhibitors. After incubating at 4 °C for 30 min, the mixture was spun at 87,920g for 30 min, and the supernatants were used as nuclear extracts.

Immunoprecipitation and western blotting. Human 293T cells (2 × 10⁶ cells) were transfected by the calcium-phosphate precipitation method with a total of 16 µg of expression plasmids for Flag-tagged or HA-tagged protein. Cells were collected 30 h after the transfection and lysed in lysis buffer containing 50 mM Tris-HCl (pH 8.0), 1% NP-40, 150 mM NaCl, 5 mM EDTA, and a cocktail of protease inhibitors and phosphatase inhibitors (Roche Diagnostics). Cell lysates were precleared by incubating at 4 °C with Sepharose CL-4B (Sigma) for 2 h, and then incubated with anti-Flag M2 affinity gel for 2 h. After washing with the lysis buffer, the proteins bound to the beads were eluted by boiling in SDS sample buffer (62.5 mM Tris-HCl (pH 6.8), 2% SDS, 10% glycerol, 5% 2-mercaptoethanol, and 0.001% bromophenol blue), separated by electrophoresis on a 10% or 10–20% polyacrylamide gel, and transferred onto a PVDF membrane. The membrane was incubated with biotin-conjugated mouse anti-Flag, or mouse anti-HA (clone 12CA5), followed by incubation with HRP-conjugated streptavidin (Roche Diagnostics). The peroxidase activity was detected by Immobilon Western ECL system (Millipore). In some case, cells were lysed with SDS sample buffer, and directly subjected to SDS-PAGE, followed by western blotting.

For analysis by native PAGE¹⁴, cells were suspended in 50 mM Tris-HCl (pH 7.5) containing 150 mM NaCl, 1 mM EDTA, 2% NP-40, cocktails of protease inhibitors and phosphatase inhibitors (Roche Diagnostics). After being briefly vortexed, the mixture was spun for 10 min at 17,600g, and the supernatant (3 µg protein) was separated by electrophoresis at 20 mA on a 7.5% polyacrylamide gel (Bio-Rad) in 62.5 mM Tris-HCl buffer (pH 6.8) containing 15% glycerol and 1% deoxycholate, and transferred to a PVDF membrane for western blotting analysis.

30. Takeshita, S., Kaji, K. & Kudo, A. Identification and characterization of the new osteoclast progenitor with macrophage phenotypes being able to differentiate into mature osteoclasts. *J. Bone Miner. Res.* **15**, 1477–1488 (2000).
31. Lam, K. M. & Hao, Q. Induction of lymphocyte agglutination and lysis by Newcastle disease virus. *Vet. Microbiol.* **15**, 49–56 (1987).
32. Shiraishi, T. *et al.* Increased cytotoxicity of soluble Fas ligand by fusing isoleucine zipper motif. *Biochem. Biophys. Res. Commun.* **322**, 197–202 (2004).
33. Hanayama, R. *et al.* Identification of a factor that links apoptotic cells to phagocytes. *Nature* **417**, 182–187 (2002).
34. Hemmi, H. *et al.* A Toll-like receptor recognizes bacterial DNA. *Nature* **408**, 740–745 (2000).
35. Ray, S. & Diamond, B. Generation of a fusion partner to sample the repertoire of splenic B cells destined for apoptosis. *Proc. Natl Acad. Sci. USA* **91**, 5548–5551 (1994).
36. Kitamura, T. *et al.* Retrovirus-mediated gene transfer and expression cloning: powerful tools in functional genomics. *Exp. Hematol.* **31**, 1007–1014 (2003).
37. Morita, S., Kojima, T. & Kitamura, T. Plat-E: an efficient and stable system for transient packaging of retroviruses. *Gene Ther.* **7**, 1063–1066 (2000).
38. Higuchi, R. in *PCR Protocols: a Guide to Methods and Applications* 177–188 (Academic, 1990).
39. Mizushima, S. & Nagata, S. pEF-BOS, a powerful mammalian expression vector. *Nucleic Acids Res.* **18**, 5322 (1990).
40. Sawasaki, T., Ogasawara, T., Morishita, R. & Endo, Y. A cell-free protein synthesis system for high-throughput proteomics. *Proc. Natl Acad. Sci. USA* **99**, 14652–14657 (2002).
41. Donella Deana, A. *et al.* An investigation of the substrate specificity of protein phosphatase 2C using synthetic peptide substrates; comparison with protein phosphatase 2A. *Biochim. Biophys. Acta* **1051**, 199–202 (1990).
42. Umeda, I. O., Nakata, H. & Nishigori, H. Identification of protein phosphatase 2C and confirmation of other protein phosphatases in the ocular lenses. *Exp. Eye Res.* **79**, 385–392 (2004).
43. Van Veldhoven, P. P. & Mannaerts, G. P. Inorganic and organic phosphate measurements in the nanomolar range. *Anal. Biochem.* **161**, 45–48 (1987).
44. Yoneyama, M. *et al.* Direct triggering of the type I interferon system by virus infection: activation of a transcription factor complex containing IRF-3 and CBP/p300. *EMBO J.* **17**, 1087–1095 (1998).

45. Fujisawa-Sehara, A., Hanaoka, K., Hayasaka, M., Hiromasa-Yagami, T. & Nabeshima, Y. Upstream region of the myogenin gene confers transcriptional activation in muscle cell lineages during mouse embryogenesis. *Biochem. Biophys. Res. Commun.* 191, 351–356 (1993).
46. Kawane, K. et al. Structure and promoter analysis of murine CAD and ICAD genes. *Cell Death Differ.* 6, 745–752 (1999).
47. Shapiro, D. J., Sharp, P. A., Wahli, W. W. & Keller, M. J. A high-efficiency HeLa cell nuclear transcription extract. *DNA* 7, 47–55 (1988).

LETTERS

Helical extension of the neuronal SNARE complex into the membrane

Alexander Stein¹, Gert Weber^{2,3}, Markus C. Wahl^{2,3} & Reinhard Jahn¹

Neurotransmission relies on synaptic vesicles fusing with the membrane of nerve cells to release their neurotransmitter content into the synaptic cleft, a process requiring the assembly of several members of the SNARE (soluble *N*-ethylmaleimide-sensitive factor attachment protein receptor) family. SNAREs represent an evolutionarily conserved protein family that mediates membrane fusion in the secretory and endocytic pathways of eukaryotic cells^{1–3}. On membrane contact, these proteins assemble *in trans* between the membranes as a bundle of four α -helices, with the energy released during assembly being thought to drive fusion^{4–6}. However, it is unclear how the energy is transferred to the membranes and whether assembly is conformationally linked to fusion. Here, we report the X-ray structure of the neuronal SNARE complex, consisting of rat syntaxin 1A, SNAP-25 and synaptobrevin 2, with the carboxy-terminal linkers and transmembrane regions at 3.4 Å resolution. The structure shows that assembly proceeds beyond the already known core SNARE complex⁷, resulting in a continuous helical bundle that is further stabilized by side-chain interactions in the linker region. Our results suggest that the final phase of SNARE assembly is directly coupled to membrane merger.

SNAREs are characterized by conserved stretches of 60–70 amino acids termed SNARE motifs^{2,8,9}. In syntaxin 1A and synaptobrevin 2, the SNARE motifs are connected by short linkers to C-terminal transmembrane regions (TMRs) (Fig. 1a). SNAP-25 is anchored to the plasma membrane by means of palmitoyl chains bound to cysteine residues in a loop region connecting its two SNARE motifs. SNARE motifs are largely unstructured, but spontaneously assemble into helical SNARE complexes^{10,11}. The X-ray structure of the synaptic core complex, which is paradigmatic for all other SNARE complexes studied so far, revealed four intertwined, parallel α -helices, with each helix being provided by a different SNARE motif. The centre of the bundle contains 16 stacked layers of interacting, mostly hydrophobic side chains⁷.

Assembly of SNARE complexes is initiated amino-terminally and proceeds towards the C terminus in a zipper-like fashion, thus pulling the membranes together^{12,13}. However, it is unclear whether zipping extends all the way into the membrane or whether the linkers between the helical SNARE motifs and the transmembrane regions remain free and flexible, with the final step in membrane merger being effected by other factors^{14–17}.

First, we tested whether the linkers and TMRs add stability to the neuronal SNARE complex. Whereas the core complex containing only the SNARE motifs is already known to be very stable¹⁸, the complex including linkers and TMRs of both syntaxin 1 and synaptobrevin 2 was significantly more resistant to thermal and chemical denaturation than the core complex alone (Fig. 1b, c). When only the TMR of syntaxin 1A was present, resistance to thermal and chemical denaturation approximated that of the core complex. When only the

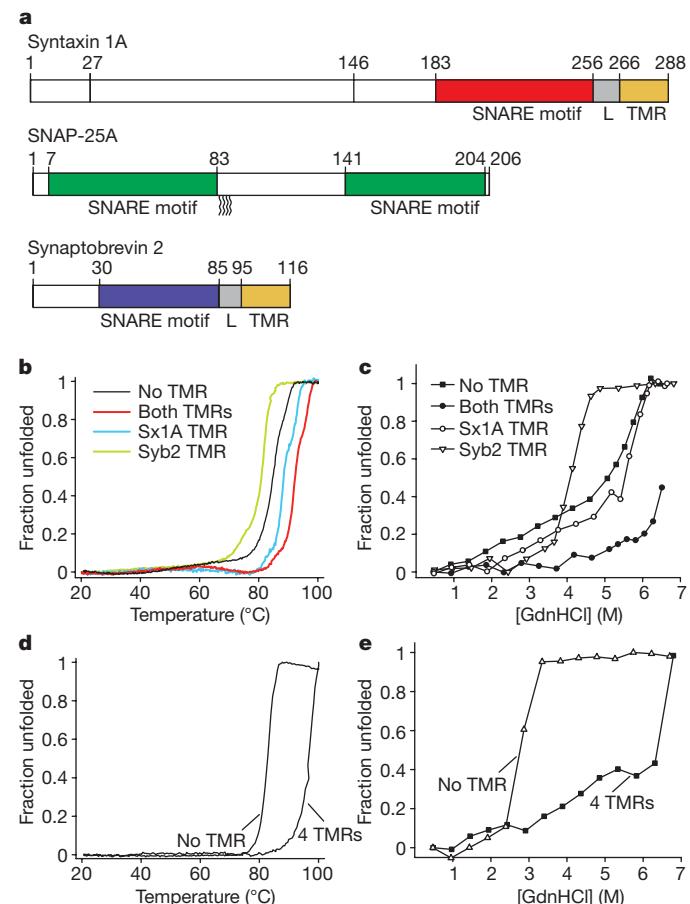


Figure 1 | Linkers and transmembrane regions add stability to SNARE complexes. **a**, Protein fragments used in this study. The complex used for crystallization contained all coloured segments. For circular dichroism (CD) measurements, complexes were formed with the same syntaxin 1A fragment, but with full-length synaptobrevin 2 and SNAP-25a (all cysteines replaced by serines). **b–e**, Unfolding of SNARE complexes, monitored by CD spectroscopy at 222 nm. **b**, Thermal unfolding of synaptic SNARE complexes in which the TMRs were either present or in which one or both of the TMRs were lacking. Note that the complex used for crystallization unfolded at approximately 97 °C (see Table 1). Sx1a, syntaxin 1A; Syb2, synaptobrevin 2. **c**, Unfolding of the same complexes at increasing concentrations of guanidine hydrochloride (GdnHCl). **d**, Thermal unfolding of endosomal SNARE complexes consisting of syntaxin 7, syntaxin 8, Vti1b and endobrevin, either containing or lacking its four TMRs. **e**, Unfolding of the endosomal complexes at increasing concentrations of GdnHCl.

¹Department of Neurobiology, ²Research Group X-ray Crystallography, Max Planck Institute for Biophysical Chemistry, 37077 Göttingen, Germany. ³Freie Universität Berlin, Fachbereich Biologie, Chemie, Pharmazie, Institut für Chemie und Biochemie, AG Strukturbiochemie, Takustraße 6, D-14195 Berlin, Germany.

TMR of synaptobrevin 2 was present, stability was lower than that of the core complex, suggesting that the presence of the free C-terminal portion of synaptobrevin interferes with the packing of the four-helix bundle further upstream. Similarly, an endosomal SNARE complex that was composed of syntaxin 7, syntaxin 8, Vti1b and endobrevin and that contained all four TMRs was more stable than the corresponding core complex (Fig. 1d, e). These findings are in agreement with an earlier study reporting increased thermal stability of the SDS-resistant band pattern in SDS-polyacrylamide gel electrophoresis and protection of the linker region from trypsin digestion, and suggest that stabilizing interactions between SNAREs extend into the linker and TMRs¹⁹.

Second, to obtain structural information about the crucial C-terminal region, we solved the X-ray structure of a neuronal SNARE complex that included the TMRs of syntaxin 1A and synaptobrevin 2. Using *n*-nonyl β-D-glucopyranoside as detergent, we obtained crystals of space group C2 that diffracted to 3.4 Å. To solve the structure, initial experimental phases were obtained by single-wavelength anomalous dispersion (SAD) phasing using a complex with selenomethionine-labelled syntaxin 1A, which was crystallized and diffracted to a resolution of 4.3 Å (see Supplementary Fig. 1a–c for densities obtained by SAD phasing and Supplementary Fig. 2a, b for final $2F_o - F_c$ maps).

As previously seen in the structure of the core complex⁷, the four SNARE motifs form a four-helix bundle. The most striking feature is that both syntaxin 1A and synaptobrevin 2 form continuous helices throughout their SNARE motifs, linker regions and TMRs (see Fig. 2a for an overview, Fig. 2b for an electrostatic surface plot, Fig. 2c for a close-up view on the linkers and TMRs, and Fig. 2d for a schematic representation of the interacting amino acids). In the crystal, hydrophobic contacts between the TMRs of synaptobrevin 2 and syntaxin 1A mediate the association of four complexes through their TMRs, forming an X-shaped assembly. Four synaptobrevin 2 TMRs build

the core of this complex, surrounded by four TMRs from syntaxin 1A (Supplementary Fig. 3a).

The two complexes in an asymmetric unit are very similar (root mean squared deviation (r.m.s.d.) = 0.6 Å). The structures of the four-helical core of the complex closely resemble the previously published structures (r.m.s.d. = 0.90 and 0.82 Å compared with the structure in ref. 20, except for the last layer (+8), which is not completely resolved). The linker region contains a collar of aromatic residues that is surrounded predominantly by basic residues (Fig. 3). Tyr 257 is part of this layer and is deeply buried in a pocket formed by three flanking lysine residues of syntaxin 1A (Lys 253, Lys 256 and Lys 260) and four residues of synaptobrevin 2 (Lys 85, Arg 86, Trp 89 and Asn 92). In addition, Lys 264 of syntaxin 1A engages in a hydrogen bond with Asn 92 of synaptobrevin 2. To investigate the importance of amino acid contacts in the linker region, we mutated several residues of syntaxin 1A to alanines (which are readily accommodated in α-helices) and subjected the corresponding complexes to thermal unfolding. Only substitution of Tyr 257 resulted in reduced thermal stability, whereas the effects of the other mutations on the melting transition temperature were not significant (Table 1). Apparently, intermolecular side-chain contacts in the aromatic layer are crucial for stability and may stiffen the linker region.

Considerable electron density is observed around the linker regions, which cannot be ascribed to protein residues, but rather seems to result from sulphate ions, glycylglycylglycine and *n*-nonyl β-D-glucopyranoside molecules in the crystallization buffer. Although owing to the limited resolution the positions of glycylglycylglycine and *n*-nonyl β-D-glucopyranoside are uncertain, the sulphate ions are in contact with multiple positively charged side chains. In a biological membrane, these basic residues would probably be stabilized by lipid head-groups. Indeed, the polybasic region in syntaxin 1A is known to interact with negatively charged lipids.

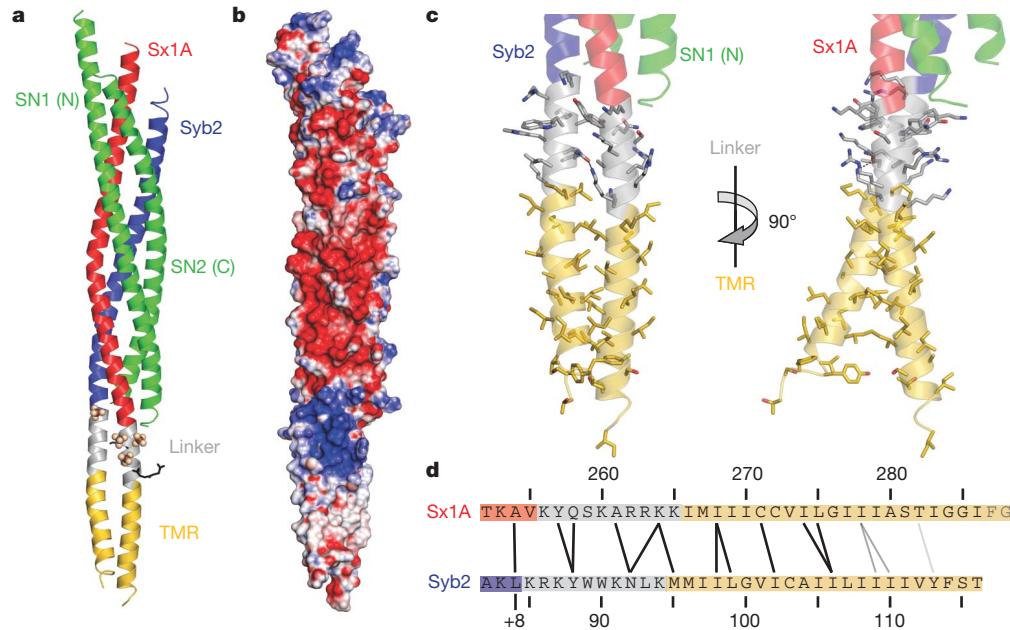


Figure 2 | Synaptobrevin 2 and syntaxin 1A form continuous helices.

a, Ribbon plot of the synaptic SNARE complex including linkers and TMRs (colour coding as in Fig. 1a). Sulphate ions and two glycylglycylglycine molecules are depicted as spheres and black sticks, respectively. b, Surface plot showing the electrostatic potential of the synaptic SNARE complex (blue, positive charge; red, negative charge). The electrostatic surface was contoured between -16 and $+16$ kT e $^{-1}$. c, Ribbon plot showing linkers and TMRs of synaptobrevin 2 and syntaxin 1A. Side chains are shown as sticks, with carbons coloured as the corresponding backbone; oxygen, red; nitrogen, blue. The hydrogen bond between Asn 92 in synaptobrevin 2 and

Lys 264 in syntaxin 1A is depicted as a black dashed line. The right panel is rotated by 90° about the vertical axis as indicated. d, Amino acids forming interactions between the two linkers and TMRs. Amino acids that are close to each other are connected by black lines, those further away but facing towards each other by grey lines. The last two amino acids in syntaxin 1A were not resolved in the structure. +8 indicates the C-terminal layer of the four-helix bundle²⁷. Note that this layer is not completely resolved as electron density for the last four amino acids of the second helix of SNAP-25 (including the layer-forming amino acid Met 202) is absent.

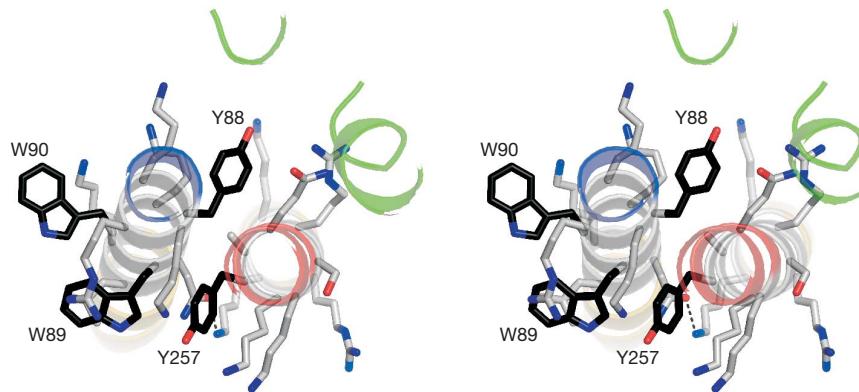


Figure 3 | An aromatic layer appears to be crucial for linker contact. Stereo view of the aromatic residues in the linker regions viewed from the N terminus. Carbon atoms in the side chains of aromatic amino acids are shown in black; otherwise, the same colour code as in Fig. 2 is used.

Mutation of these residues reduced the affinity for these lipids and resulted in a decrease in evoked secretion from bovine chromaffin cells²¹.

C-terminally adjacent to the linker regions, the TMRs remain in contact, with residues Met 95, Ile 98, Leu 99, Ile 102 and Ile 106 in synaptobrevin 2, and Ile 268, Cys 271, Ile 274 and Leu 275 in syntaxin 1A largely contributing to the interaction surface. Beyond this region, the two TMRs veer away from each other. We also solved the structure of a complex purified and crystallized using *n*-heptyl β-D-glucopyranoside as a detergent. These crystals diffracted to 4.8 Å resolution, belong to space group $I2_12_12_1$, and again harbour two very similar complexes per asymmetric unit (see Supplementary Fig. 2c for the model overlaid on the $F_o - F_c$ map). Although the overall structures of the complexes are very similar to the structures of the complexes in the monoclinic form ($r.m.s.d. = 0.9 \text{ \AA}$), the C-terminal parts of the synaptobrevin 2 TMRs deviate slightly, indicating some conformational flexibility (see Supplementary Fig. 4a, b). In the $I2_12_12_1$ crystals, again four SNARE complexes form an X-shaped assembly, with the contacts being mainly mediated by the TMRs. However, the overall packing differs from C2 packing (see Supplementary Fig. 3b). Although both inter-complex contacts at the TMRs and the angle between the TMRs are likely to be determined by crystal packing, it is conceivable that such contacts have a role during the transition states of the fusion reaction.

Helical continuity throughout the linkers and TMRs is unexpected, considering that perturbation of the linkers between the SNARE motifs and the TMRs, although adversely affecting fusion, seem to be better tolerated than perturbations of the SNARE motifs, in which single point mutations can abolish function^{12,22–24}. Insertion of short flexible elements between SNARE motif and TMR has only mild effects, but fusion of both biological and artificial membranes is

progressively inhibited when the length of the insert is extended. Recently, synaptobrevin 2 variants carrying linker mutations were overexpressed in chromaffin cells that lack endogenous synaptobrevins and hence Ca^{2+} -regulated exocytosis²⁵. Whereas wild-type synaptobrevin 2 fully rescued exocytosis, insertion of six amino acids just in front of the TMR of synaptobrevin 2 almost completely abolished fast secretion. Furthermore, although the Ca^{2+} threshold for release remained unchanged, the delay between the rise of intracellular Ca^{2+} concentration and the onset of release was increased, indicating a ‘weakening’ of the fusion machine operating downstream of the Ca^{2+} trigger. On the other hand, the effect of inserting two supposedly helix-breaking proline residues in the linker on exocytosis

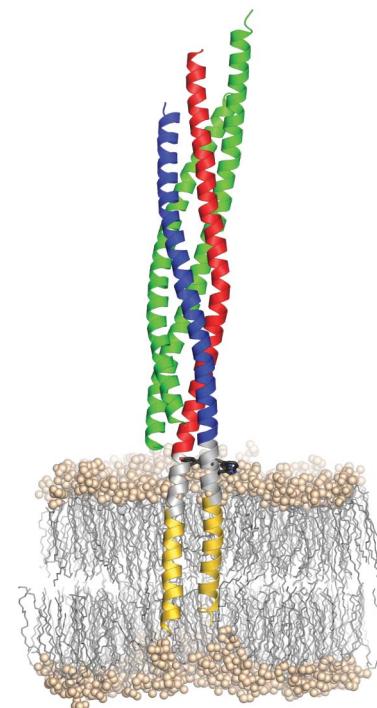


Figure 4 | Model of the synaptic SNARE complex inserted in a membrane. As a landmark, aromatic residues (black sticks) within the linker region (grey) are shown. The hydrophilic head groups of the phospholipids are shown as balls, their aliphatic chains as sticks. The position of the complex in the palmitoyl-oleoyl-phosphatidylethanolamine (POPE) bilayer was estimated from a short (54.9 ns) molecular dynamics simulation, where the apolar parts of the TMRs were initially centred within the hydrophobic part of the bilayer. Because PE head groups are highly abundant in animal membranes and PO is a relatively short tail group, POPE lipids were chosen to mimic a simple membrane with a thickness of approximately 4.5–5.0 nm.

Table 1 | Melting temperatures for all neuronal SNARE complexes

Protein constructs	<i>n</i>	Melting temperature (°C)	s.d.
Wild type	9	94.9	1.4
K256A	4	93.2	0.5
Y257A	5	90.6	0.3
Q258A	4	95.1	0.7
K260A	4	94.4	0.4
R262A	4	93.5	0.8
R263A	6	94.2	0.6
K264A	5	94.4	0.6
K265A	4	94.2	0.4
No TMR	5	84.9	0.6
Syb2 TMR	4	80.7	0.7
Sx1A TMR	5	88.6	0.6
Used in crystallization	4	96.5	0.5

For investigating the role of single amino acids in complex stability, complexes containing both TMRs and the indicated mutations in syntaxin 1A were prepared. Ellipticity at 222 nm was monitored and the inflection point of the melting transition taken to be the melting temperature.

was surprisingly mild, in agreement with an earlier study in which no serious growth defects were observed when two prolines were inserted into the linker of the yeast syntaxin homologue Sso1p²⁶. However, the mild phenotypes may be at least partially explained by overexpression of the mutant proteins, because a much stronger phenotype was observed when the (Pro)₂-Sso1p variant was driven by the endogenous Sso1 promoter²⁶.

Figure 4 shows a model in which the SNARE complex was placed into a lipid bilayer. It represents the final stage of SNARE-mediated membrane fusion in which both TMRs come to rest in the same bilayer (*cis* complex). However, it raises interesting questions about the structural features of the complex before zippering is completed (*trans* complex). Although it is conceivable that in a partially zippered complex the non-assembled regions between the N-terminally assembled helical parts and the (presumably) helical transmembrane domains are unstructured, it is also possible that small helical regions exist that may provide a binding site for late regulatory proteins such as synaptotagmin and complexin. The model illustrates that the four-helix bundle of the SNARE motifs sits as a perpendicular rod directly on top of the phospholipid bilayer. With the linkers of synaptobrevin 2 and syntaxin 1A buried in the top layer of the membrane, it is difficult to imagine how the four-helix bundle can be fully zippered before membrane merger is at least initiated, but we cannot exclude that zippering is confined to SNARE complexes surrounding the neck of a fusion pore that may form only transiently (kiss-and-run mode). Accordingly, it is understandable that structural perturbations in the linkers are less disruptive for fusion and may be at least partially compensated by engaging more SNARE complexes in the fusion reaction. Our data strongly support the view that SNAREs constitute a robust and simple fusion catalyst. SNAREs operate as nanomachines in which the energy barrier of membrane fusion is overcome by the energy provided during progressive N–C zippering of the SNARE proteins all the way into the membrane.

METHODS SUMMARY

All SNARE monomers were separately expressed from pET28a as His₆-tagged proteins in *Escherichia coli* BL21 (DE3) cells and purified by Ni²⁺-NTA and ion exchange chromatography. SNARE complexes were then assembled from monomers and purified by ion exchange and size-exclusion chromatography. Diffraction data were collected on beamline X10SA at the Swiss Light Source of the Paul Scherrer Institut (Switzerland) and processed with HKL-2000. Initial phases were obtained from SAD data from crystals containing selenomethionine-labelled syntaxin 1A that diffracted to 4.3 Å resolution. For the final model building, native diffraction data to 3.4 Å resolution were used. Model building and refinement were performed using the programs COOT and PHENIX, respectively.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 26 January; accepted 15 May 2009.

Published online 1 July 2009.

- Martens, S. & McMahon, H. T. Mechanisms of membrane fusion: disparate players and common principles. *Nature Rev. Mol. Cell Biol.* **9**, 543–556 (2008).
- Jahn, R. & Scheller, R. H. SNAREs—engines for membrane fusion. *Nature Rev. Mol. Cell Biol.* **7**, 631–643 (2006).
- Rizo, J. & Rosenmund, C. Synaptic vesicle fusion. *Nature Struct. Mol. Biol.* **15**, 665–674 (2008).
- Pelham, H. R., Banfield, D. K. & Lewis, M. J. SNAREs involved in traffic through the Golgi complex. *Cold Spring Harb. Symp. Quant. Biol.* **60**, 105–111 (1995).
- Hanson, P. I., Heuser, J. E. & Jahn, R. Neurotransmitter release — four years of SNARE complexes. *Curr. Opin. Neurobiol.* **7**, 310–315 (1997).

- Lin, R. C. & Scheller, R. H. Structural organization of the synaptic exocytosis core complex. *Neuron* **19**, 1087–1094 (1997).
- Sutton, R. B., Fasshauer, D., Jahn, R. & Brunger, A. T. Crystal structure of a SNARE complex involved in synaptic exocytosis at 2.4 Å resolution. *Nature* **395**, 347–353 (1998).
- Brunger, A. T. Structure and function of SNARE and SNARE-interacting proteins. *Q. Rev. Biophys.* **38**, 1–47 (2005).
- Kloepper, T. H., Kienle, N. C. & Fasshauer, D. An elaborate classification of SNARE proteins sheds light on the conservation of the eukaryotic endomembrane system. *Mol. Biol. Cell* **18**, 3463–3471 (2007).
- Fasshauer, D. et al. A structural change occurs upon binding of syntaxin to SNAP-25. *J. Biol. Chem.* **272**, 4582–4590 (1997).
- Fasshauer, D. et al. Structural changes are associated with soluble N-ethylmaleimide-sensitive fusion protein attachment protein receptor complex formation. *J. Biol. Chem.* **272**, 28036–28041 (1997).
- Sorensen, J. B. et al. Sequential N-to C-terminal SNARE complex assembly drives priming and fusion of secretory vesicles. *EMBO J.* **25**, 955–966 (2006).
- Pobbati, A. V., Stein, A. & Fasshauer, D. N- to C-terminal SNARE complex assembly promotes rapid membrane fusion. *Science* **313**, 673–676 (2006).
- Su, Z., Ishitsuka, Y., Ha, T. & Shin, Y. K. The SNARE complex from yeast is partially unstructured on the membrane. *Structure* **16**, 1138–1146 (2008).
- Dennison, S. M., Bowen, M. E., Brunger, A. T. & Lenz, B. R. Neuronal SNAREs do not trigger fusion between synthetic membranes but do promote PEG-mediated membrane fusion. *Biophys. J.* **90**, 1661–1675 (2006).
- Jackson, M. B. & Chapman, E. R. Fusion pores and fusion machines in Ca²⁺-triggered exocytosis. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 135–160 (2006).
- Rizo, J., Chen, X. & Arac, D. Unraveling the mechanisms of synaptotagmin and SNARE function in neurotransmitter release. *Trends Cell Biol.* **16**, 339–350 (2006).
- Fasshauer, D., Antonin, W., Subramaniam, V. & Jahn, R. SNARE assembly and disassembly exhibit a pronounced hysteresis. *Nature Struct. Biol.* **9**, 144–151 (2002).
- Poirier, M. A. et al. Protease resistance of syntaxin·SNAP-25·VAMP complexes. Implications for assembly and structure. *J. Biol. Chem.* **273**, 11370–11377 (1998).
- Ernst, J. A. & Brunger, A. T. High resolution structure, stability, and synaptotagmin binding of a truncated neuronal SNARE complex. *J. Biol. Chem.* **278**, 8630–8636 (2003).
- Lam, A. D. et al. SNARE-catalyzed fusion events are regulated by syntaxin1A-lipid interactions. *Mol. Biol. Cell* **19**, 485–497 (2008).
- McNew, J. A. et al. The length of the flexible SNAREpin juxtamembrane region is a critical determinant of SNARE-dependent fusion. *Mol. Cell* **4**, 415–421 (1999).
- Wang, Y., Dulubova, I., Rizo, J. & Sudhof, T. C. Functional analysis of conserved structural elements in yeast syntaxin Vam3p. *J. Biol. Chem.* **276**, 28598–28605 (2001).
- Deak, F., Shin, O. H., Kavalali, E. T. & Sudhof, T. C. Structural determinants of synaptobrevin 2 function in synaptic vesicle fusion. *J. Neurosci.* **26**, 6668–6676 (2006).
- Kesavan, J., Borisovska, M. & Bruns, D. v-SNARE actions during Ca²⁺-triggered exocytosis. *Cell* **131**, 351–363 (2007).
- Van Komen, J. S. et al. The polybasic juxtamembrane region of Sso1p is required for SNARE function *in vivo*. *Eukaryot. Cell* **4**, 2017–2028 (2005).
- Fasshauer, D., Sutton, R. B., Brunger, A. T. & Jahn, R. Conserved structural features of the synaptic fusion complex: SNARE proteins reclassified as Q- and R-SNAREs. *Proc. Natl. Acad. Sci. USA* **95**, 15781–15786 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank U. Ries for technical assistance, C. Kutzner and H. Grubmüller for providing the model shown in Fig. 4, R. Lührmann for access to his X-ray facilities, D. Fasshauer and N. Pavlos for comments on the manuscript, W. Antonin for providing the expression constructs of endosomal SNAREs and the staff of beamlines PX1 and PX2 of the Swiss Light Source for support during diffraction data collection.

Author Contributions A.S. and R.J. planned the project, A.S. performed the experiments, A.S., G.W. and M.C.W. analysed the data and A.S. and R.J. wrote the manuscript. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to R.J. (rjahn@gwdg.de).

METHODS

Protein constructs. The SNARE proteins syntaxin 1A, 7 and 8, SNAP-25A, synaptobrevin 2 and endobrevin from rat and Vti1b from mouse were used. The following expression constructs (all in pET28a) have been described: syntaxin 1A (residues 180–262 and 183–288)^{28,29}, syntaxin 7 (residues 159–236), syntaxin 8 (residues 136–213), SNAP-25A (residues 7–83, 141–204 and a full-length version in which all cysteines were mutated to serines)^{30,31}, synaptobrevin 2 (residues 1–96 and 1–116)^{28,29}, Vti1b (residues 130–206) and endobrevin (residues 1–74 and 1–100)^{31–33}. For crystallization of the neuronal SNARE complex, DNA encoding a synaptobrevin 2 fragment lacking the N-terminal proline-rich region was cloned into pET28a via NdeI/XbaI restriction sites (residues 30–116). Genes of the endosomal SNAREs containing TMRs were also cloned into pET28a via NdeI/XbaI restriction sites: syntaxin 7 (residues 159–261), syntaxin 8 (residues 136–236), Vti1b (residues 130–232).

Protein expression and purification. Protein purification was carried out essentially as described^{13,29,30,32,34}. Briefly, all proteins were expressed in *E. coli* strain BL21 (DE3) and purified by Ni²⁺-NTA affinity chromatography followed by ion exchange chromatography on an Äkta system (GE Healthcare). Syntaxin 1A (residues 183–288) and syntaxin 7 (residues 159–261) were desalted after Ni²⁺-NTA affinity chromatography and removal of their tags on a HiPrep 26/10 Desalting column (GE Healthcare) equilibrated with 20 mM Tris buffer pH 8.8, 500 mM NaCl, 50 mM *n*-octyl β-D-glucopyranoside, 1 mM tris-(2-carboxyethyl) phosphine (TCEP) without further purification by ion-exchange chromatography. Synaptobrevin 2 (residues 1–116 and 30–116) and all endosomal SNAREs containing TMRs were purified by ion exchange chromatography in the presence of 50 mM *n*-octyl β-D-glucopyranoside. Selenomethionine-labelled syntaxin 1A was expressed in minimal medium containing selenomethionine as described³⁵. Purification was performed in the presence of 10 mM β-mercaptoethanol during Ni²⁺-NTA affinity chromatography, and 10 mM dithiothreitol in all subsequent steps. All SNARE complexes were assembled by overnight incubation of the monomers at 4 °C, using syntaxin 1A or syntaxin 7 as the limiting components. Complexes were then separated from free monomers by ion-exchange chromatography in the presence of 50 mM *n*-octyl β-D-glucopyranoside (neuronal complexes used for CD experiments), 12 mM *n*-nonyl β-D-glucopyranoside, 140 mM *n*-heptyl β-D-glucopyranoside (complexes used for crystallization) or 0.03% (w/v) *n*-dodecyl β-D-maltopyranoside (endosomal complex). Complexes used for crystallization were additionally purified by size-exclusion chromatography using a Superdex 200 column equilibrated with 10 mM Tris, pH 8.8, 1 M NaCl, 1 mM TCEP and 12 mM *n*-nonyl β-D-glucopyranoside or 140 mM *n*-heptyl β-D-glucopyranoside. **Crystallization and data collection.** For crystallization, the neuronal SNARE complex was reconstituted from the following monomers: syntaxin 1A (residues 183–288), SNAP-25A (residues 7–83 and 141–204) and synaptobrevin 2 (residues 30–116). Initial crystals diffracting to about 8 Å resolution were obtained using *n*-octyl β-D-glucopyranoside as a stabilizing detergent. Variation of the length of the aliphatic chain by ±1 methylene group yielded crystals that diffracted to about 4.8 and 4.3 Å for *n*-heptyl β-D-glucopyranoside and *n*-nonyl β-D-glucopyranoside, respectively. The complex purified in the presence of *n*-nonyl β-D-glucopyranoside crystallized in sitting drops at 20 °C. Crystals were obtained by mixing protein at a concentration of 15–20 mg ml⁻¹ in a 1:1 ratio with reservoir solution containing 100 mM HEPES, pH 7.3, 200 mM Li₂SO₄, 100 mM NaCl, 29% PEG 400 and 30 mM Gly-Gly-Gly. Crystals appeared after 2–3 days and grew within 2 weeks to about 0.1–0.3 mm in each dimension. Before freezing, single crystals were transferred to a 5 μl drop of 100 mM HEPES, pH 7.3, 200 mM Li₂SO₄, 100 mM NaCl, 35% PEG 400, 30 mM Gly-Gly-Gly, 12 mM *n*-nonyl β-D-glucopyranoside and placed on ice for 1–3 h. Crystals were then flash-frozen in a 100 K nitrogen stream. This procedure improved diffraction to a minimum Bragg spacing of 3.4 Å.

For the crystallization of the complex purified in the presence of *n*-heptyl β-D-glucopyranoside, the reservoir solution contained 100 mM Tris, pH 8.8, 50 mM MgCl₂, 16% 2-methyl-2,4-pentanediol and 3% (w/v) sorbitol. These crystals were cryo-protected by transfer into perfluoropolyether and then flash-frozen in liquid nitrogen. Data sets were collected on frozen crystals at 100 K on the X10SA beamline at the Swiss Light Source of the Paul Scherrer Institut (Switzerland) using a MAR225 CCD detector. The data were indexed, integrated and scaled with the HKL-2000 package³⁶ (Supplementary Table 1).

Structure determination. The crystals obtained with *n*-nonyl β-D-glucopyranoside belong to space group C2 ($a = 265.4 \text{ \AA}$, $b = 135.5 \text{ \AA}$, $c = 58.7 \text{ \AA}$, $\beta = 96.3^\circ$) with two complexes and 82% solvent in the crystallographic asymmetric unit. Initial experimental phases were obtained by SAD phasing using a complex with selenomethionine-labelled syntaxin 1A, which was crystallized and diffracted to a resolution of 4.3 Å. Using the SHELX package³⁷ 12 out of 14 methionines present in a crystallographic asymmetric unit were located. An initial model was generated

by placing a high-resolution structure of the neuronal core complex²⁰ into the experimental electron density using MOLREP³⁸ and manually extending the structure into the linkers and TMRs with COOT³⁹. This model was transferred to a 3.4 Å native data set and further refined by manual model building and automated refinement with REFMAC5⁴⁰ including TLS (translation, liberation and screw-rotation) refinement⁴¹. Final refinement was performed using the PHENIX.REFINE program⁴². During all refinement cycles, non-crystallographically independent complexes. The positions of methionines in syntaxin 1A, including one methionine in the TMR, were subsequently verified using the anomalous difference density from the selenomethionine data set.

The structure in the orthorhombic form was solved by molecular replacement using the structure obtained for the monoclinic form as a search model. The shift in the TMR was clearly observed as difference electron density in the $F_o - F_c$ map and corrected by manual model building followed by automated refinement using the PHENIX.REFINE program, imposing strong restraints on the non-crystallographic symmetry and geometry, and refining one B factor per amino acid⁴². TLS refinement did not lead to a considerable improvement of R-factors and was therefore not performed.

CD spectroscopy. CD measurements were performed using a Chirascan instrument (Applied Photophysics). Hellma quartz cuvettes with a path length of 0.1 cm were used. For thermal denaturation experiments, complexes were dissolved to 5 μM in 20 mM sodium phosphate, pH 7.4, 500 mM NaF, 0.1 mM TCEP, 34 mM *n*-octyl β-D-glucopyranoside. The ellipticity at 222 nm was recorded between 20 and 100 °C with a temperature increment of 15 °C h⁻¹ for the neuronal complexes, and 30 °C h⁻¹ for the endosomal complexes. The temperature was measured with a temperature probe placed in the solution. For chemical denaturation experiments, complexes were dissolved to 3 μM in 20 mM sodium phosphate, 0.02% *n*-dodecyl β-D-maltopyranoside and the indicated amounts of GndHCl. After at least 8 h incubation in the respective buffer, far-ultraviolet CD spectra were recorded at 20 °C using steps of 0.2 nm with an averaging time of 1 s. For data analysis, ellipticity recorded at 222 nm was converted to fraction of unfolded protein. In the case of the neuronal complex with both TMRs, total unfolding was not achieved at 20 °C and approximately 6.5 M GndHCl. The solution was therefore heated to 95 °C and the corresponding ellipticity at 222 nm used for normalization. For the other complexes this procedure did not result in any further change in ellipticity.

Molecular dynamics simulation. Individual lipids (229) were taken from an equilibrated POPE membrane simulation and subsequently placed around the complex such that the average area per lipid molecule was that of the donor membrane. The remaining unoccupied volume of the simulation box was then filled with 26,629 SPC/E (simple point charge/extended) water molecules⁴³. For the SNARE complex the optimized potentials for liquid simulations all-atom force field was used⁴⁴ and lipid parameters were taken from ref. 45. After energy-minimizing the SNARE complex, water and membrane were separately equilibrated for 10 ps each. For the subsequent simulation the protein was kept as a rigid body, whereas phospholipids and water were allowed to move freely. The simulation was performed with GROMACS 4.0.3 (ref. 46). Cutoffs were set to 1.0 nm, long-range electrostatics were calculated with the particle-mesh Ewald method. Berendsen temperature and pressure coupling was used to keep the system at 300 K and 1 bar⁴⁷. Using a time step length of 2 fs, the LINCS and SETTLE constraint algorithms were applied^{48,49}.

28. Fasshauer, D., Eliason, W. K., Brunger, A. T. & Jahn, R. Identification of a minimal core of the synaptic SNARE complex sufficient for reversible assembly and disassembly. *Biochemistry* **37**, 10354–10362 (1998).
29. Schuette, C. G. et al. Determinants of liposome fusion mediated by synaptic SNARE proteins. *Proc. Natl. Acad. Sci. USA* **101**, 2858–2863 (2004).
30. Pobbat, A. V. et al. Structural basis for the inhibitory role of tomosyn in exocytosis. *J. Biol. Chem.* **279**, 47192–47200 (2004).
31. Fasshauer, D. et al. Mixed and non-cognate SNARE complexes. Characterization of assembly and biophysical properties. *J. Biol. Chem.* **274**, 15440–15446 (1999).
32. Antonin, W. et al. A SNARE complex mediating fusion of late endosomes defines conserved properties of SNARE structure and function. *EMBO J.* **19**, 6453–6464 (2000).
33. Brandhorst, D. et al. Homotypic fusion of early endosomes: SNAREs do not determine fusion specificity. *Proc. Natl. Acad. Sci. USA* **103**, 2701–2706 (2006).
34. Stein, A. et al. Synaptotagmin activates membrane fusion through a Ca²⁺-dependent trans interaction with phospholipids. *Nature Struct. Mol. Biol.* **14**, 904–911 (2007).
35. Van Duyne, G. D. et al. Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J. Mol. Biol.* **229**, 105–124 (1993).
36. Otwowski, Z. & Minor, W. Processing of X-ray diffraction data collected in the oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
37. Sheldrick, G. M. A short history of SHELX. *Acta Crystallogr. A* **64**, 112–122 (2008).
38. Vagin, A. & Tepljakov, A. An approach to multi-copy search in molecular replacement. *Acta Crystallogr. D* **56**, 1622–1624 (2000).

39. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
40. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
41. Winn, M. D., Isupov, M. N. & Murshudov, G. N. Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr. D* **57**, 122–133 (2001).
42. Adams, P. D. et al. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58**, 1948–1954 (2002).
43. Berendsen, H. J. C., Grigera, J. R. & Straatsma, T. P. The missing term in effective pair potentials. *J. Phys. Chem.* **91**, 6269–6271 (1987).
44. Kaminski, G. A., Friesner, R. A., Tirado-Rives, J. & Jorgensen, W. L. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B* **105**, 6474–6487 (2001).
45. Berger, O., Edholm, O. & Jahnig, F. Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature. *Biophys. J.* **72**, 2002–2013 (1997).
46. Hess, B., Kutzner, C., van der Spoel, D. & Lindahl, E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **4**, 435–447 (2008).
47. Berendsen, H. J. C. et al. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690 (1984).
48. Hess, B. P-LINCS: A parallel linear constraint solver for molecular simulation. *J. Chem. Theory Comput.* **4**, 116–122 (2008).
49. Miyamoto, S. & Kollman, P. A. SETTLE: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **13**, 952–962 (1992).

LETTERS

Modulation of microRNA processing by p53

Hiroshi I. Suzuki¹, Kaoru Yamagata^{2,3}, Koichi Sugimoto⁴, Takashi Iwamoto⁵, Shigeaki Kato^{2,3} & Kohei Miyazono¹

MicroRNAs (miRNAs) have emerged as key post-transcriptional regulators of gene expression, involved in diverse physiological and pathological processes. Although miRNAs can function as both tumour suppressors and oncogenes in tumour development¹, a widespread downregulation of miRNAs is commonly observed in human cancers and promotes cellular transformation and tumorigenesis^{2–4}. This indicates an inherent significance of small RNAs in tumour suppression. However, the connection between tumour suppressor networks and miRNA biogenesis machineries has not been investigated in depth. Here we show that a central tumour suppressor, p53, enhances the post-transcriptional maturation of several miRNAs with growth-suppressive function, including miR-16-1, miR-143 and miR-145, in response to DNA damage. In HCT116 cells and human diploid fibroblasts, p53 interacts with the Drosha processing complex through the association with DEAD-box RNA helicase p68 (also known as DDX5) and facilitates the processing of primary miRNAs to precursor miRNAs. We also found that transcriptionally inactive p53 mutants interfere with a functional assembly between Drosha complex and p68, leading to attenuation of miRNA processing activity. These findings suggest that transcription-independent modulation of miRNA biogenesis is intrinsically embedded in a tumour suppressive program governed by p53. Our study reveals a previously unrecognized function of p53 in miRNA processing, which may underlie key aspects of cancer biology.

In mammalian miRNA biogenesis, the primary transcripts of miRNA genes (pri-miRNAs) are cleaved into hairpin intermediates (pre-miRNAs) by the nuclear RNase III Drosha and further processed to mature miRNAs by cytosolic Dicer, another RNase-III-related enzyme⁵. The Drosha complex comprises Drosha, the DiGeorge syndrome critical region gene 8 (DGCR8) and multiple RNA-associated proteins including the DEAD box RNA helicases p68 and p72 (the latter also known as DDX17)⁶. p68 and p72 are required for the maturation of some, but not all, miRNAs⁷.

A widespread decrease in mature miRNAs is often observed in various human malignancies^{2,8–10}. In addition to genomic and epigenetic alterations^{9,11}, the deregulation might be attributable to the impairment of miRNA-processing steps^{12,13}. Although several reports have described altered expression of miRNA-processing factors in tumour tissues¹⁴, these phenomena may not be a widespread genetic event. We therefore postulate a direct connection between tumour suppressor networks and miRNA biogenesis machineries. We first compared miRNAs induced by DNA damage and/or p53 activation^{15–17}, and also p68/p72-regulated miRNAs at the processing steps⁷, and noticed that p72-dependent miRNAs tend to be upregulated under DNA-damaging conditions (Supplementary Fig. 1). This observation led us to test the involvement of p53 in miRNA processing, in accord with a previous report of a functional interaction between p53 and p68/p72 (ref. 18).

To identify miRNAs potentially regulated by p53, we measured the expression levels of a selected group of miRNAs by quantitative polymerase chain reaction with reverse transcription (qRT-PCR) after treatment with the DNA-damaging agent doxorubicin, a potent p53 inducer, in the *p53* wild-type HCT116 human colon cancer cell line. Consistent with previous findings^{15–17} was our observation that exposure to doxorubicin increased not only a known p53 target, miR-34a, but also a set of mature miRNAs such as miR-15a, miR-16-1, miR-23a, miR-26a, miR-103, miR-143, miR-145, miR-203 and miR-206 (Fig. 1a). To determine whether their accumulation is mediated transcriptionally or post-transcriptionally, we measured the expression levels of primary precursor, precursor and mature forms of these miRNAs. miR-34a, a direct transcriptional target of p53, increased at all primary, precursor and mature miRNA levels. In contrast, several other miRNAs, including miR-16-1, miR-143, miR-145 and miR-206, showed no significant change in primary transcripts but did show increased expression of precursor and mature miRNAs; this was confirmed by northern blot analysis (Fig. 1a and Supplementary Fig. 2). Similar results were obtained in two human diploid cell lines, WI-38 and TIG-3, and TOV21G ovarian cancer cells (Supplementary Figs 3 and 4, respectively).

The DNA-damage-induced upregulation of pre-miRNAs and mature miRNAs was diminished in *p53*^{-/-} HCT116 cells (Fig. 1b). siRNA-mediated p53 knockdown also attenuated these increments of pre-miRNAs and mature miRNAs in p53-competent WI-38 cells and HCT116 cells (Fig. 1c and Supplementary Fig. 5). Further, p68 or p72 knockdown abolished the increase of these miRNAs at the precursor and mature levels (Fig. 1c and Supplementary Fig. 5), indicating that several miRNAs are post-transcriptionally upregulated in a p53-dependent and p68/p72-dependent manner after DNA damage. The miRNAs analysed above have been shown to be decreased in various human cancers^{1,8,10,13,19,20}, and they decreased the proliferation rate of HCT116 cells (Supplementary Fig. 6a). We further found that novel targets of these miRNAs include important regulators of the cell cycle and cell proliferation, such as *K-Ras* (as a target of miR-143) and *CDK6* (as a target of miR-16, miR-26a, miR-107, miR-145 and miR-206) (Supplementary Fig. 6b). Each miRNA suppressed the expression of luciferase reporter gene fused to the fragments of 3' untranslated regions (3' UTRs) of *K-Ras* and *CDK6*, containing the corresponding potential target sites (Supplementary Figs 7 and 8). Moreover, luciferase activity was significantly decreased in response to DNA damage (Supplementary Fig. 9). Thus, p53/p68-dependent regulation of several miRNAs might support a growth inhibitory function of p53.

We next investigated the molecular interaction between the Drosha complex and p53, because p53 has previously been shown to interact with p68 (ref. 18). Ectopically expressed Drosha and p53 were associated in HCT116 cells (Fig. 2a). We also found that p53 accumulated and interacted with endogenous Drosha/DGCR8 on

¹Department of Molecular Pathology, Graduate School of Medicine, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. ²Institute of Molecular and Cellular Biosciences, University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-0032, Japan. ³ERATO, Japan Science and Technology Agency, 4-1-8 Honcho, Kawaguchi, Saitama 332-0012, Japan. ⁴Division of Hematology, Department of Internal Medicine, Juntendo University School of Medicine, 2-1-1 Hongo, Bunkyo-ku, Tokyo 113-8421, Japan. ⁵The Center for Education in Laboratory Animal Research and Department of Biomedical Sciences, College of Life and Health Sciences, Chubu University, 1200 Matsumoto-cho, Kasugai, Aichi 487-8501, Japan.

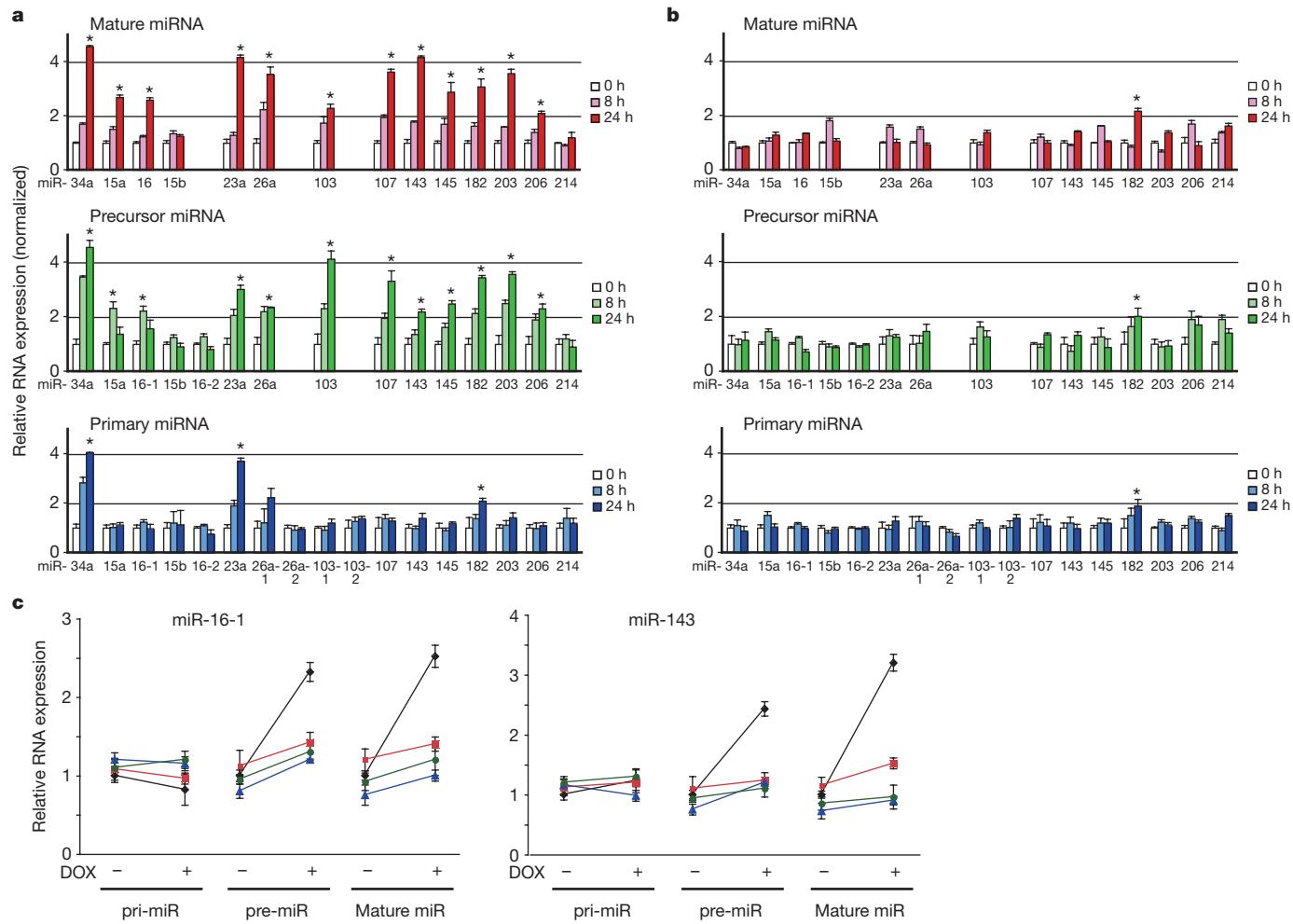


Figure 1 | Post-transcriptional modification of miRNA biogenesis by p53 and p68/p72 in response to DNA damage. **a, b**, The expression levels of primary precursor, precursor and mature forms of the indicated miRNAs were examined in *p53* wild-type (*p53*^{WT}) HCT116 cells (**a**) and *p53*^{-/-} HCT116 cells (**b**) after treatment for 8 or 24 h with 0.2 μ g ml⁻¹ doxorubicin (asterisk, $P < 0.05$ compared with no treatment; $n = 3$). Error bars represent s.e.m. **c**, Effects of p53, p68 or p72 knockdown on miRNA elevation after treatment with doxorubicin (DOX). WI-38 human diploid fibroblasts were transfected with control siRNA (black) or siRNA for *p53* (red), *p68* (blue) or *p72* (green). After treatment with doxorubicin, the expression levels of primary, precursor and mature miRNAs were compared by qRT-PCR analysis.

treatment with doxorubicin in HCT116, WI-38, TIG-3 and MCF7 cells (Fig. 2b and Supplementary Fig. 10). The association of Drosha and p53 was markedly decreased by treatment with RNase A1 and on

s.e.m. **c**, Effects of p53, p68 or p72 knockdown on miRNA elevation after treatment with doxorubicin (DOX). WI-38 human diploid fibroblasts were transfected with control siRNA (black) or siRNA for *p53* (red), *p68* (blue) or *p72* (green). After treatment with doxorubicin, the expression levels of primary, precursor and mature miRNAs were compared by qRT-PCR analysis.

p68/p72 knockdown by siRNA (Fig. 2c, d), indicating that p53 interacts with the Drosha complex through certain RNA species and the *p68/p72* RNA helicase(s). In addition, we observed that p53 associates with Drosha or *p68* through a carboxy-terminal half of the central DNA-binding domain (Supplementary Figs 11 and 12).

In the process of DNA-damage-mediated induction of pre-miRNAs and mature miRNAs, the transcription rates of pri-miR-16-1 and miR-143 were constant in nuclear run-on assay (Fig. 3a) and the levels of unprocessed pri-miRNA tended to decrease (Supplementary Fig. 13). Therefore, on the basis of an interaction between Drosha and p53, we addressed whether p53 modulates the *in vivo* association

Figure 2 | Association between p53 and Drosha complex. **a, b**, Immunoprecipitation assays were performed to detect interaction between ectopic (**a**) and endogenous (**b**) p53 and Drosha in HCT116 cells. An association between endogenous p53 and Drosha was observed under DNA-damaging conditions (**b**). ALY served as a negative control for the immunoprecipitation assay. Ab, antibody; DOX, doxorubicin; IB, immunoblot; IP, immunoprecipitation. **c**, RNA dependence of interaction between p53 and Drosha. Immunoprecipitates of Drosha prepared from HCT116 cells treated for 8 h with doxorubicin were digested with RNase V1 (nuclease for double-stranded RNA and double-stranded DNA) or RNase A1 (RNase for single-stranded RNA), separated to supernatant (sup.) and beads, and subjected to immunoblot analysis. **d**, Immunoblot analysis of Drosha immunoprecipitates from HCT116 cells transfected with control or *p68* and *p72* siRNA, followed by treatment for 8 h with doxorubicin.

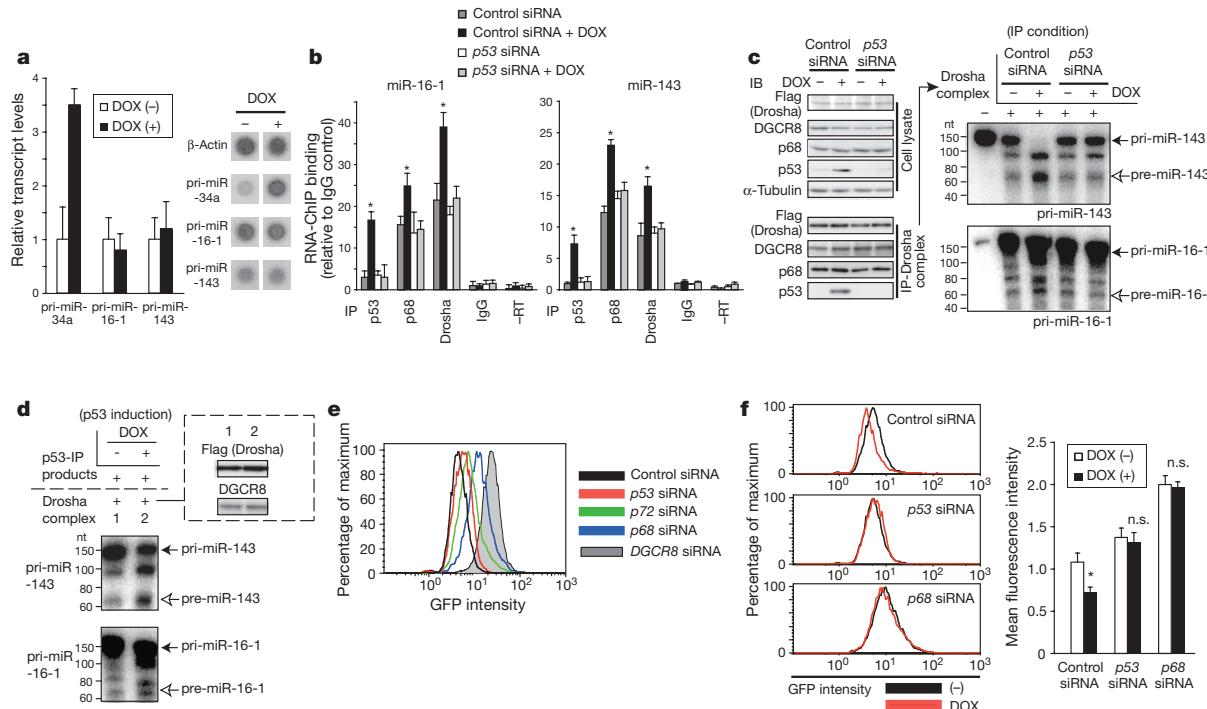


Figure 3 | p53 facilitates Drosha-mediated pri-miRNA processing.

a, Nuclear run-on assay. After treatment with doxorubicin (DOX), the transcription rates of pri-miR-34a, miR-16-1 and miR-143 were measured and normalized against β-actin levels ($n = 3$). Representative dot-blots are shown in the right panel. **b**, RNA-ChIP analysis for association between pri-miR-16-1/pri-miR-143 and p53, p68 or Drosha in WI-38 cells. After transfection with control or p53 siRNA, WI-38 cells were treated for 8 h with or without doxorubicin, and endogenous p53, p68 or Drosha were immunoprecipitated and subjected to RT-PCR analysis. As controls, RNA samples untreated with reverse transcriptase (–RT) or immunoprecipitated with non-specific IgG (IgG) were subjected to PCR (asterisk, $P < 0.05$ compared with control siRNA (DOX (–)); $n = 3$). **c**, In vitro pri-miRNA processing assay of pri-miR-16-1/pri-miR-143 with immunoprecipitated Flag-Drosha complex from HCT116 cells after siRNA transfection and

treatment with doxorubicin. Filled and open arrows indicate the precursor and processed RNAs, respectively. The left panel shows an immunoblot of whole cell lysates and immunoprecipitated Flag-Drosha complex of each sample used for the *in vitro* processing assay. nt, nucleotides. **d**, Addition of immunoprecipitated p53 from HCT116 cells treated for 8 h with doxorubicin facilitated the cleavage of pri-miR-16-1/pri-miR-143 by the Flag-Drosha complex. The top right panel shows an immunoblot of immunoprecipitated Drosha. **e**, **f**, Fluorescence-activated cell sorting analysis of HCT116 cells carrying fusion genes encoding the EGFP cDNA and pri-miR-143/pri-miR-145 (PMXs-puro-EGFP-miR-145/miR-143). DGCR8 or p68 knockdown strongly augmented GFP signals (**e**). Changes in GFP signals were measured after treatment with doxorubicin under p53 or p68 knockdown (**f**). Fluorescence intensities are shown as means and s.e.m.

of pri-miRNAs with the Drosha complex and Drosha function. RNA-chromatin immunoprecipitation (ChIP) analysis showed that p53 associates with pri-miR-16-1/pri-miR-143 after treatment with doxorubicin (Fig. 3b and Supplementary Fig. 14). Treatment with doxorubicin also significantly enhanced the association between these pri-miRNAs and Drosha or p68 in a p53-dependent manner, in both WI-38 cells and HCT116 cells (Fig. 3b and Supplementary Fig. 14), suggesting that the p53 pathway promotes the recruitment of the p68-Drosha complex to the target pri-miRNAs.

We also performed an *in vitro* pri-miRNA processing assay by incubating radiolabelled pri-miR-16-1/pri-miR-143 substrate with immunoprecipitated Flag-tagged Drosha complex from HCT116 cells treated with or without doxorubicin. Drosha complex from doxorubicin-treated cells contained higher pri-miRNA processing activity than that from mock-treated cells, and the higher processing activity was abolished by knockdown of p53 (Fig. 3c). Furthermore, an *in vitro* combination of immunoprecipitated p53 and Drosha complex reconstructed the facilitation of pri-miRNA processing (Fig. 3d). We next established *in vivo* cellular systems for convenient and real-time monitoring of Drosha function²¹. HCT116 cells were infected with retroviral constructs carrying fusion genes encoding enhanced green fluorescent protein (EGFP) and a short segment of pri-miR-143/pri-miR-145. We predicted that the EGFP transcripts should lose their poly(A) tail by Drosha-mediated cleavage of pri-miRNAs, resulting in their poor stability and decreased translation. In this cell line, GFP intensity was inversely correlated with Drosha activity (Supplementary Fig. 15). DGCR8 or p68 knockdown

strongly enhanced the GFP signal, whereas p72 knockdown modestly enhanced it (Fig. 3e). In this system we observed that DNA damage by doxorubicin caused a decrease in GFP signal; this attenuation was decreased by knockdown of p53 or p68 (Fig. 3f). Taken together, these results demonstrate that p53 enhances Drosha-mediated processing of several miRNAs both *in vitro* and *in vivo* under DNA-damage-inducing conditions.

p53 mutations are frequently observed in human tumours, and most of these have missense mutations that lead to a loss of wild-type p53 transcriptional activity and to the accumulation of stable p53 mutants in cancer cells²². Some mutant p53 proteins such as R175H and R273H have been shown to exert oncogenic activities in addition to a loss of function^{23–25}. We therefore examined the effects of several tumour-derived transcriptionally inactive p53 mutants (C135Y, R175H and R273H) on miRNA processing through their introduction into p53-null HCT116 cells. The p53 mutants suppressed the precursor and mature miRNA levels of miR-16-1, miR-143 and miR-206, in comparison with the constant level of pri-miRNAs, whereas wild-type p53 increased the pre-miRNA and mature miRNA expression levels of these miRNAs (Fig. 4a–c). The mutant p53 also decreased the production of mature and precursor miR-16-1 and miR-143 from the ectopically expressed pri-miR-16-1/pri-miR-143 (Supplementary Figs 16 and 17), suggesting that the mutant p53 hinders miRNA processing in a transcription-independent manner. We verified that the p53 mutants heightened the GFP signals as well as p68 knockdown in the *in vivo* monitoring of microprocessor activity (Fig. 4d and Supplementary Fig. 18). To elucidate

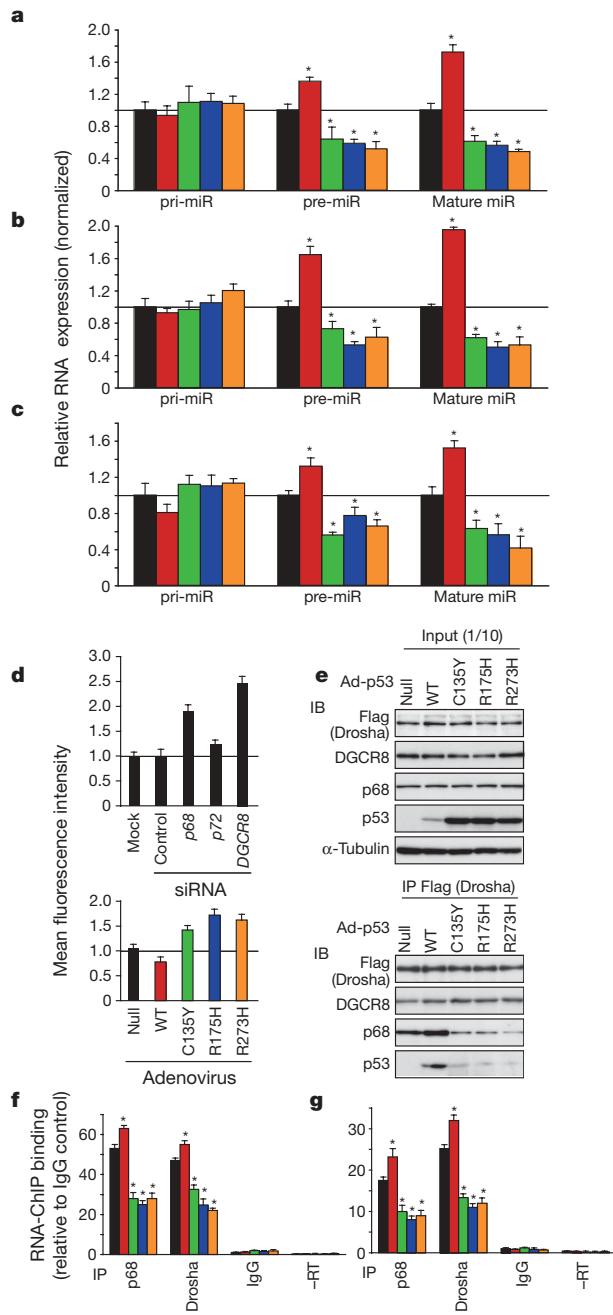


Figure 4 | Deregulation of miRNA processing by mutant p53. **a–c**, *p53*^{−/−} HCT116 cells were infected with adenovirus carrying wild-type (red) or tumour-derived transcriptionally inactive mutant p53 (C135Y (green), R175H (blue) and R273H (orange)) or control adenovirus (Ad-null; black). The amounts of primary precursor, precursor and mature forms of miR-16-1 (**a**), miR-143 (**b**) and miR-206 (**c**) were examined (asterisk, $P < 0.05$ compared with Ad-null; $n = 3$). Error bars represent s.e.m. **d**, Fluorescence-activated cell sorting analysis of *p53*^{−/−} HCT116 cells harbouring pMXs-puro-EGFP-miR-145/miR-143. Mean fluorescence intensities of GFP signals were measured after infection with Ad-p53 (wild-type (WT), C135Y, R175H and R273H) or transfection with siRNAs. Error bars represent s.e.m. **e**, Immunoprecipitation (IP) assays were performed after infection with Ad-p53 (WT, C135Y, R175H and R273H) or Ad-null. IB, immunoblotting. **f, g**, RNA-ChIP analysis. After infection of *p53*^{−/−} HCT116 cells with Ad-p53, endogenous proteins were immunoprecipitated by anti-p68 or anti-Drosha antibodies and subjected to RT-PCR analysis with miR-16-1 (**f**) or miR-143 (**g**) primers (asterisk, $P < 0.05$ compared with Ad-null; $n = 3$). Colour coding is as in **a–c**.

the mechanisms of decreased miRNA processing by the p53 mutants, we examined the effect of mutant p53 expression on the association between Drosha and p68. As shown in Fig. 4e, the accumulation of

mutant p53 decreased the interaction between Drosha and p68, but p53 mutants still interacted with p68 (data not shown). In contrast, wild-type p53 expression induced a modest increase in the Drosha–p68 association (Fig. 4e). Finally, RNA-ChIP analysis also revealed that the mutant p53 decreased the association between pri-miR-16-1/pri-miR-143 and p68 or Drosha (Fig. 4f, g). Tumour-derived p53 mutations might therefore endow some different and opposite properties on this protein, leading to interference with miRNA biogenesis by negative titration of Drosha microprocessor components and interference with Drosha/p68 accessibility to several pri-miRNAs. In support of this notion, widespread downregulation of miRNA was prominent in late-stage or high-grade ovarian cancers with a higher frequency of p53 mutation and overexpression⁹.

This study raises a possibility that p53 functions as a global modifier of gene expression through the regulation of miRNA processing. The production of mRNA isoforms with a different 3' UTR, a major docking platform for miRNAs, is regulated by alternative polyadenylation. It has recently been shown that increased proliferation is coupled with widespread decrease in lengths and miRNA target site abundance in 3' UTRs of mRNA²⁶. Consistently, avoidance of strict gene regulation by miRNAs might be a general feature of cancer cells. Thus, a miRNA-controlled gene regulatory layer may itself serve as a part of global program for cellular proliferation and have an intrinsic tumour-suppressive function in conjunction with the p53 network. In addition, a loss of mature miRNAs has been shown to result in increased DNA damage and p53 activity, indicating a reciprocal connection between the p53 and miRNA pathways²⁷.

Apart from the important role of p53 as a sequence-specific transcription factor, previous studies have revealed that its transcription-independent roles contribute to the regulation of apoptosis and to DNA repair and recombination^{28,29}. Here we have shown an unexpected function of p53 in miRNA processing. Further investigation would provide insights into how great a portion of pri-miRNAs are regulated by p53 and how the specificities of recognition of pri-miRNAs are determined. A recent study also demonstrated a similar regulatory pathway of miRNA biosynthesis mediated by nuclear SMAD proteins³⁰, indicating intimate crosstalks between miRNA-processing machineries and nuclear factors. The present study offers an insight into a new approach to cancer treatment by modulating the p53 and miRNA pathways.

METHODS SUMMARY

Cell lines. The human colon cancer cell line HCT116 and isogenic *p53*^{−/−} HCT116 cells were provided by B. Vogelstein. Two human diploid fibroblast lines, WI-38 and TIG-3, were obtained from the RIKEN Cell Bank and the Health Science Research Resources Bank, respectively. The TOV21G human ovarian cancer cell line was purchased from the American Type Culture Collection.

qRT-PCR analysis. qRT-PCR assays were performed for measurement of the expression levels of primary, precursor and mature miRNAs as described previously³⁰.

RNA-ChIP. RNA-ChIP was performed as described previously⁷.

In vitro pri-miRNA processing analysis. The *in vitro* pri-miRNA processing assay was performed as described previously^{6,7}. *In vitro* transcription was conducted with linearized pcDNA3 vector containing pri-miRNA. The processing reaction, containing the immunoprecipitated Drosha complex (with/without immunoprecipitated p53) and the internally [α -³²P]UTP-labelled pri-miRNA was incubated at 37 °C for 90 min.

In vivo monitoring of pri-miRNA processing. Retroviral vectors with fusion genes of the EGFP cDNA and pri-miRNA²¹ were introduced into HCT116 cells. High-responder clones to DGCR8 knockdown were selected. The change in GFP signals was measured by flow cytometric analysis.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 17 April; accepted 4 June 2009.

- Esquela-Kerscher, A. & Slack, F. J. Oncomirs—microRNAs with a role in cancer. *Nature Rev. Cancer* 6, 259–269 (2006).
- Lu, J. et al. MicroRNA expression profiles classify human cancers. *Nature* 435, 834–838 (2005).

3. Kumar, M. S., Lu, J., Mercer, K. L., Golub, T. R. & Jacks, T. Impaired microRNA processing enhances cellular transformation and tumorigenesis. *Nature Genet.* **39**, 673–677 (2007).
4. Chang, T. C. et al. Widespread microRNA repression by Myc contributes to tumorigenesis. *Nature Genet.* **40**, 43–50 (2008).
5. Kim, V. N. MicroRNA biogenesis: coordinated cropping and dicing. *Nature Rev. Mol. Cell Biol.* **6**, 376–385 (2005).
6. Gregory, R. I. et al. The Microprocessor complex mediates the genesis of microRNAs. *Nature* **432**, 235–240 (2004).
7. Fukuda, T. et al. DEAD-box RNA helicase subunits of the Drosha complex are required for processing of rRNA and a subset of microRNAs. *Nature Cell Biol.* **9**, 604–611 (2007).
8. Ozen, M., Creighton, C. J., Ozdemir, M. & Ittmann, M. Widespread deregulation of microRNA expression in human prostate cancer. *Oncogene* **27**, 1788–1793 (2008).
9. Zhang, L. et al. Genomic and epigenetic alterations deregulate microRNA expression in human epithelial ovarian cancer. *Proc. Natl Acad. Sci. USA* **105**, 7004–7009 (2008).
10. Marton, S. et al. Small RNAs analysis in CLL reveals a deregulation of miRNA expression and novel miRNA candidates of putative relevance in CLL pathogenesis. *Leukemia* **22**, 330–338 (2008).
11. Calin, G. A. et al. Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proc. Natl Acad. Sci. USA* **101**, 2999–3004 (2004).
12. Thomson, J. M. et al. Extensive post-transcriptional regulation of microRNAs and its implications for cancer. *Genes Dev.* **20**, 2202–2207 (2006).
13. Michael, M. Z., O'Connor, S. M., van Holst Pellekaan, N. G., Young, G. P. & James, R. J. Reduced accumulation of specific microRNAs in colorectal neoplasia. *Mol. Cancer Res.* **1**, 882–891 (2003).
14. Karube, Y. et al. Reduced expression of Dicer associated with poor prognosis in lung cancer patients. *Cancer Sci.* **96**, 111–115 (2005).
15. He, L. et al. A microRNA component of the p53 tumour suppressor network. *Nature* **447**, 1130–1134 (2007).
16. Chang, T. C. et al. Transactivation of miR-34a by p53 broadly influences gene expression and promotes apoptosis. *Mol. Cell* **26**, 745–752 (2007).
17. Tarasov, V. et al. Differential regulation of microRNAs by p53 revealed by massively parallel sequencing: miR-34a is a p53 target that induces apoptosis and G1-arrest. *Cell Cycle* **6**, 1586–1593 (2007).
18. Bates, G. J. et al. The DEAD box protein p68: a novel transcriptional coactivator of the p53 tumour suppressor. *EMBO J.* **24**, 543–553 (2005).
19. Bueno, M. J. et al. Genetic and epigenetic silencing of microRNA-203 enhances ABL1 and BCR-ABL1 oncogene expression. *Cancer Cell* **13**, 496–506 (2008).
20. Kondo, N., Toyama, T., Sugiura, H., Fujii, Y. & Yamashita, H. miR-206 expression is down-regulated in estrogen receptor α -positive human breast cancer. *Cancer Res.* **68**, 5004–5008 (2008).
21. Tsutsui, M. et al. Establishment of cells to monitor Microprocessor through fusion genes of microRNA and GFP. *Biochem. Biophys. Res. Commun.* **372**, 856–861 (2008).
22. Soussi, T. & Beroud, C. Assessing TP53 status in human tumours to evaluate clinical outcome. *Nature Rev. Cancer* **1**, 233–240 (2001).
23. Soussi, T. p53 alterations in human cancer: more questions than answers. *Oncogene* **26**, 2145–2156 (2007).
24. Song, H. & Xu, Y. Gain of function of p53 cancer mutants in disrupting critical DNA damage response pathways. *Cell Cycle* **6**, 1570–1573 (2007).
25. Adorno, M. et al. A mutant-p53/Smad complex opposes p63 to empower TGF β -induced metastasis. *Cell* **137**, 87–98 (2009).
26. Sandberg, R., Neilson, J. R., Sarma, A., Sharp, P. A. & Burge, C. B. Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. *Science* **320**, 1643–1647 (2008).
27. Mudhasani, R. et al. Loss of miRNA biogenesis induces p19Arf-p53 signaling and senescence in primary cells. *J. Cell Biol.* **181**, 1055–1063 (2008).
28. Chipuk, J. E. et al. Direct activation of Bax by p53 mediates mitochondrial membrane permeabilization and apoptosis. *Science* **303**, 1010–1014 (2004).
29. Sengupta, S. & Harris, C. C. p53: traffic cop at the crossroads of DNA repair and recombination. *Nature Rev. Mol. Cell Biol.* **6**, 44–55 (2005).
30. Davis, B. N., Hilyard, A. C., Lagna, G. & Hata, A. SMAD proteins control DROSHA-mediated microRNA maturation. *Nature* **454**, 56–61 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank H. Matsuyama, K. Kiyono, S. Ehata and R. A. Saito for discussion; M. Saitoh, K. Miyazawa, T. Watabe, K. Horiguchi, T. Shirakihara, K. Harada, M. Oka, A. Mizutani, T. Yamazaki and Y. Yoshimatsu for technical advice and reagents; T. Yokochi and Y. Morishita for encouragement; and all members of the Department of Molecular Pathology, University of Tokyo for assistance. This work was supported by KAKENHI (Grant-in-Aid for Scientific Research no. 17016011) on priority areas 'New strategies for cancer therapy based on advancement of basic research' and the Global Center of Excellence Program for 'Integrative Life Science Based on the Study of Biosignaling Mechanisms' from the Ministry of Education, Culture, Sports, Science and Technology of Japan. H.I.S. is supported by a research fellowship of the Japan Society for the Promotion of Science for Young Scientists.

Author Contributions H.I.S. conceived and designed the research, performed experiments and analyses and wrote the paper. Y.K., T.I. and S.K. provided key materials. K.S. and K.M. supervised the whole project and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to K.M. (miyazono@m.u-tokyo.ac.jp).

METHODS

Antibodies. The following antibodies were used: Flag epitope tag M2 (Sigma); Myc epitope tag 9E10 (Pharmingen) or 71D10 (no. 2278) (Cell Signaling), p53 DO-1 (sc-126) or FL-393 (sc-6243) (Santa Cruz), p68 05-850 (Upstate), p72 ab24601 (Abcam), CDK6 B-10 (sc-7961) (Santa Cruz), K-Ras F234 (sc-30) (Santa Cruz), α -tubulin DM-1A (Sigma), DGCR8 10996-1-AP (Proteintech), Drosha 07-717 (Upstate) and ALY 11G5 (sc-32311) (Santa Cruz).

qRT-PCR assays and northern blot analysis. qRT-PCR assays were performed for measurement of the expression levels of primary, precursor and mature miRNAs, as described previously³⁰. In brief, total RNA was extracted with Trizol (Invitrogen) and subjected to reverse transcription with the Quantitect Reverse Transcription kit (Qiagen) according to the manufacturer's instructions. qRT-PCR was performed with the 7500 Fast Real-Time PCR System (Applied Biosystems). For detection of mature miRNAs, TaqMan MicroRNA assay kit (Applied Biosystems) was used in accordance with the manufacturer's protocol. Data analysis was performed by using the comparative C_t method. Results were normalized to β -actin or U6 snRNA. The expression ratio of U6 snRNA and β -actin were constant on treatment with doxorubicin. For measurement of the precursor miRNA expression, as an alternative approach, small RNA fraction was purified with the miRvana miRNA isolation kit (Ambion) and subjected to reverse transcription and qRT-PCR analyses (normalization by U6 snRNA); similar results were obtained by this method. The primer sequences used are given in the Supplementary Information. Northern blot analysis was performed as described previously^{6,7}. Oligonucleotides complementary to miRNAs were end-labelled with [γ -³²P]ATP and used as probes for northern analysis.

siRNAs. The siRNAs (except that for DGCR8) and control siRNA (ON-TARGET plus non-targeting pool) were purchased from Dharmacon (SMARTpool). siRNA for DGCR8 and control siRNA were purchased from Invitrogen (Stealth RNAi). siRNAs (20 nM) were introduced into cultured cells with HiPerfect transfection reagent (Qiagen). The transfected cells were used for subsequent experiments after 24–48 h. The target sequences of these siRNA duplexes are given in Supplementary Information.

Construction of p53 deletion vectors. p53 deletion constructs (Δ TAD (Δ 1–45), Δ PRD (Δ 62–91), Δ DBD (Δ 114–290), Δ DBD-N (Δ 114–212), Δ DBD-C (Δ 213–290) and Δ C (Δ 291–393)) were prepared as described in previous reports^{31–34}.

Adenovirus. Wild-type and mutant p53 (C135Y, R175H and R273H) cDNAs were cloned into a pENTR vector (Invitrogen) and introduced into the pAd/CMV/V5-DEST vector by using LR Clonase (Invitrogen). HEK293A cells were transfected with pAd/CMV/p53 (WT/C135Y/R175H/R273H) after linearization with *PacI*. Viral particles were isolated by three freeze-thaw cycles and amplified by reinfection to HEK293A cells.

Lentivirus. HCT116 cells stably expressing Flag-Drosha were generated by lentiviral infection system (a gift from H. Miyoshi). Flag-Drosha cDNA was transferred into lentivirus vector (CSII-CMV-RfA) by means of a pENTR vector, with LR clonase. The lentivirus production was carried by transfection of HEK293FT cells with vector construct, pCMV-VSV-G-RSV-Rev and pCAG-HIVgp. The culture supernatants were collected 48 h after transfection, and viral particles were introduced to HCT116 cells.

Immunoprecipitation and immunoblot assay. Cells were lysed with a buffer containing 1% Nonidet P40, 20 mM Tris-HCl pH 7.4, 150 mM NaCl, 5 mM EDTA and 1% protease inhibitor mixture (Nacalai Tesque). For immunoprecipitation, cleared lysates were incubated with the indicated antibodies. For RNase V1 (nuclease for dsRNA/dsDNA) and RNase A1 (RNase for ssRNA) digestion, the immunoprecipitated proteins were washed twice with RNase digestion buffer (Ambion) and suspended in the same buffer containing RNase V1 or A1 (Ambion). The samples were incubated at 37 °C for 30 min and quickly separated by centrifugation into two fractions: beads containing the still-bound proteins and supernatant containing the released proteins. Total cell lysates or proteins in immunoprecipitates were subjected to SDS-PAGE and transferred to Fluoro Trans W membrane (Pall). Immunoblotting was performed with the indicated antibodies.

Proliferation and cell cycle analysis. Pre-miR miRNA Precursors were purchased from Ambion. For growth experiments, cells were seeded in 96-well plates and transfected with pre-miR precursors (20 nM), and growth rates were measured with a WST-8 assay kit (Nacalai Tesque). The cell cycle distribution was determined with a FACScan flow cytometer (Beckman Coulter) using CycleTEST PLUS DNA Reagent kit (BD Biosciences).

UTR reporter assay. The 3' UTR segments of *K-Ras* and *CDK6* were cloned into the Psicheck 2 dual luciferase reporter vector (Promega). HCT116 cells were then

transfected with each reporter construct with and without pre-miR precursors. Cell extracts were prepared 24–48 h after transfection, and the ratio of *Renilla* to firefly luciferase was measured with the Dual-Luciferase Reporter Assay System (Promega). The primer sequences used for cloning of the 3' UTR segments are given in the Supplementary Information.

RNA-ChIP. RNA-ChIP was performed as described previously^{7,35}. All the buffers used in this study contained 0.5 U μ l⁻¹ RNase inhibitor (Toyobo). Nuclei from HCT116 cells were first isolated from 1% formaldehyde-fixed cells and used for chromatin fragmentation. After immunoprecipitation with anti-p53, anti-p68 or anti-Drosha antibody, washing and elution, the precipitated RNA-DNA pellets were resuspended in 100 μ l of nuclease-free water containing 1 μ l of 40 U μ l⁻¹ RNase inhibitor, 5 μ l of 1 M Tris-HCl pH 7.5, 20 μ l of 50 mM MgCl₂ and 4 μ l of 10 U μ l⁻¹ DNase I. The mixture was incubated for 30 min at 37 °C and extracted once with phenol/chloroform (5:1). RNA was precipitated with ethanol and dissolved in 30 μ l of nuclease-free water. A 29- μ l aliquot of the RNA was used for a 60- μ l cDNA synthesis reaction and qRT-PCR analysis. The primer sequences used are given in Supplementary Information.

Nuclear run-on assay. The nuclear run-on assay was performed as described previously³⁶. An aliquot (0.5 μ g) of each PCR fragment amplified by primers for the qRT-PCR assay was dot-blotted onto positively charged nylon membranes (Roche). After isolation of the nucleus and run-on transcription reaction, RNA probe was prepared with Trizol reagent. Membranes were probed with ³²P-labelled run-on RNA, and then analysed with a phosphorimager.

In vitro pri-miRNA processing analysis. *In vitro* pri-miRNA processing assay was performed as described previously^{6,7,37}. *In vitro* transcription was performed with linearized pcDNA3 vector containing pri-miR-16-1, pri-miR-143 or pri-miR-145. The processing reaction, containing the immunoprecipitated Drosha complex from HCT116 cells (with or without immunoprecipitated p53 from HCT116 cells), 3 μ l of a solution containing 32 mM MgCl₂, 10 mM ATP, 200 mM creatine phosphate, 20 U μ l⁻¹ RNase inhibitor, internally [α -³²P]UTP-labelled pri-miRNA and buffer (20 mM Tris-HCl pH 7.9, 0.1 M KCl, 10% glycerol, 5 mM dithiothreitol, 0.2 mM phenylmethylsulphonyl fluoride) was added to a final volume of 30 μ l. The reaction mixture was incubated at 37 °C for 90 min and extracted with phenol/chloroform, then with chloroform, and precipitated with glycogen and ethanol. The precipitated RNA was loaded on 15% denaturing polyacrylamide gels, and then autoradiographed.

In vivo monitoring of pri-miRNA processing. Retroviral constructs with fusion genes of the EGFP cDNA and pri-miRNA (pMXs-puro-miR-143-IRES-EGFP, pMXs-puro-EGFP-miR-145/miR-143 and pMXs-puro-chimera-miR-143-EGFP) were described previously²¹. A retroviral supernatant was obtained by transfection of G3T-hi cells by using a Retrovirus Constructive Kit Amphi (Takara Bio) in accordance with the manufacturer's instructions. Retroviral infection was performed with standard procedures, and the cells were selected with 1 μ g ml⁻¹ puromycin, 48 h after infection. High-responder clones to DGCR8 knockdown were selected. After treatment with doxorubicin, transfection with siRNA or infection with adenovirus, cells were analysed with a FACScan flow cytometer (Beckman Coulter) and the data were analysed with Flowjo (Tree Star).

Statistical analysis. Statistical differences were determined by analysis of variance followed by post-hoc comparison with the Fisher's protected least-significant-difference test. Statistical significance is shown as $P < 0.05$ (asterisk).

31. Zhao, B. X. et al. p53 mediates the negative regulation of MDM2 by orphan receptor TR3. *EMBO J.* **25**, 5703–5715 (2006).
32. Liu, G., Xia, T. & Chen, X. The activation domains, the proline-rich domain, and the C-terminal basic domain in p53 are necessary for acetylation of histones on the proximal p21 promoter and interaction with p300/CREB-binding protein. *J. Biol. Chem.* **278**, 17557–17565 (2003).
33. Roth, J., Koch, P., Contente, A. & Dobbelstein, M. Tumor-derived mutations within the DNA-binding domain of p53 that phenotypically resemble the deletion of the proline-rich domain. *Oncogene* **19**, 1834–1842 (2000).
34. Rui, Y. et al. Axin stimulates p53 functions by activation of HIPK2 kinase through multimeric complex formation. *EMBO J.* **23**, 4583–4594 (2004).
35. Ni, J. Q., Liu, L. P., Hess, D., Rietdorf, J. & Sun, F. L. *Drosophila* ribosomal proteins are associated with linker histone H1 and suppress gene transcription. *Genes Dev.* **20**, 1959–1973 (2006).
36. Kim, H. K., Lee, Y. S., Sivaprasad, U., Malhotra, A. & Dutta, A. Muscle-specific microRNA miR-206 promotes muscle differentiation. *J. Cell Biol.* **174**, 677–687 (2006).
37. Lee, Y. et al. The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**, 415–419 (2003).

LETTERS

The CREB coactivator CRTC2 links hepatic ER stress and fasting gluconeogenesis

Yiguo Wang¹, Liliana Vera¹, Wolfgang H. Fischer¹ & Marc Montminy¹

In fasted mammals, circulating pancreatic glucagon stimulates hepatic gluconeogenesis in part through the CREB regulated transcription coactivator 2 (CRTC2, also referred to as TORC2)^{1,2}. Hepatic glucose production is increased in obesity, reflecting chronic increases in endoplasmic reticulum (ER) stress that promote insulin resistance³. Whether ER stress also modulates the gluconeogenic program directly, however, is unclear. Here we show that CRTC2 functions as a dual sensor for ER stress and fasting signals. Acute increases in ER stress triggered the dephosphorylation and nuclear entry of CRTC2, which in turn promoted the expression of ER quality control genes through an association with activating transcription factor 6 alpha (ATF6 α , also known as ATF6)—an integral branch of the unfolded protein response^{4–9}. In addition to mediating CRTC2 recruitment to ER stress inducible promoters, ATF6 α also reduced hepatic glucose output by disrupting the CREB–CRTC2 interaction and thereby inhibiting CRTC2 occupancy over gluconeogenic genes. Conversely, hepatic glucose output was upregulated when hepatic ATF6 α protein amounts were reduced, either by RNA interference (RNAi)-mediated knockdown or as a result of persistent stress in obesity. Because ATF6 α overexpression in the livers of obese mice reversed CRTC2 effects on the gluconeogenic program and lowered hepatic glucose output, our results demonstrate how cross-talk between ER stress and fasting pathways at the level of a transcriptional coactivator contributes to glucose homeostasis.

Obesity is a central risk factor in the development of insulin resistance, which is characterized by an inability of insulin to inhibit glucose output from the liver and to increase glucose uptake into skeletal muscle^{10,11}. Although the underlying mechanism is unclear, obesity has been shown to disrupt insulin signalling in the liver and adipose through chronic increases in ER stress³. Because hepatic glucose production is also increased in obesity, we investigated whether ER stress signals modulate the gluconeogenic program directly.

Previous studies showing an important role for the CREB coactivator CRTC2 in promoting hepatic gluconeogenesis^{1,2} led us to examine the effects of ER stress in this setting. Exposure of primary hepatocytes to the ER stress activators thapsigargin (THA) or tunicamycin (TUN)⁶ stimulated CRTC2 dephosphorylation and nuclear entry (Fig. 1a, b and Supplementary Fig. 1); these effects were blocked when cells were pre-treated with cyclosporine, an inhibitor of the Ser/Thr phosphatase calcineurin (PPP3CA), which has been shown to mediate CRTC2 dephosphorylation^{12,13}.

We used an adenovirally encoded cAMP-responsive luciferase (Ad-CRE-luc) reporter to monitor CREB–CRTC2 activity in primary hepatocytes. Although they stimulated an ER stress-inducible reporter (X-box binding protein 1 (XBP1)-luc)^{7,14}, THA and TUN inhibited Ad-CRE-luc activity, even when cells were co-stimulated with the cAMP activator forskolin (FSK; Fig. 1a, bottom).

Having seen that ER stress promotes CRTC2 activation but not CREB-dependent transcription, we considered the potential involvement of a CRTC2 inhibitor in this process. In proteomic studies to identify cellular proteins that associate with CRTC2, we recovered the basic leucine zipper (bZIP) transcription factor ATF6 α from immunoprecipitates of endogenous CRTC2 (Supplementary Fig. 2a). We confirmed the CRTC2–ATF6 α interaction in co-immunoprecipitation studies of primary hepatocytes using endogenous and epitope-tagged proteins (Fig. 1c and Supplementary Fig. 2a).

Localized to the ER under basal conditions, ATF6 α undergoes intramembrane proteolysis and nuclear shuttling in response to ER stress, when it promotes cell viability by stimulating ER quality control

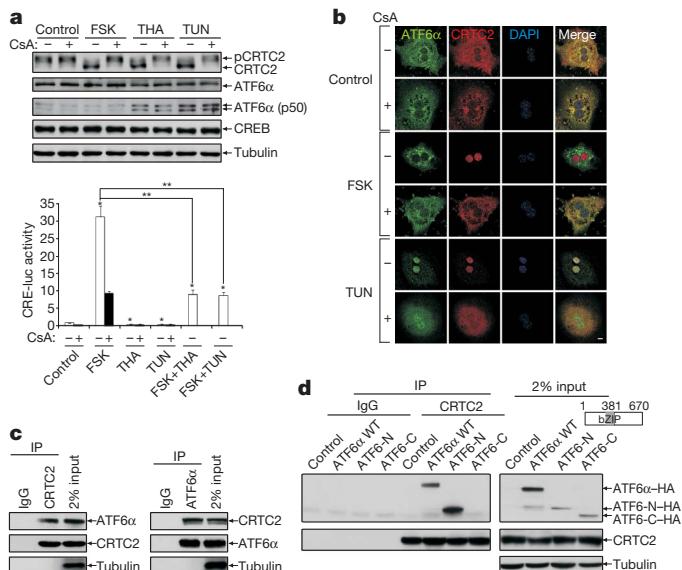


Figure 1 | Nuclear translocation and association of CRTC2 with ATF6 α in response to ER stress. **a**, Effects of ER stress activators (THA, TUN) and FSK on CRTC2 dephosphorylation (top) and Ad-CRE-luc reporter activity (bottom) in primary hepatocytes. pCRTC2, phosphorylated CRTC2. Pre-treatment with the calcineurin inhibitor cyclosporine (CsA) is indicated. The results are the average of three independent experiments; error bars denote s.e.m.; *P < 0.01, **P < 0.001. **b**, Immunostaining for CRTC2 and ATF6 α in primary hepatocytes exposed to FSK or TUN. Scale bar, 5 μ m. DAPI, 4,6-diamidino-2-phenylindole. **c**, Immunoblot showing co-immunoprecipitation (co-IP) of endogenous CRTC2 with ATF6 α in primary hepatocyte extracts. **d**, Immunoblot showing relative association of CRTC2 with the transcriptionally active ATF6-N terminal domain (amino acids 1–381; ATF6-N) or C-terminal ER luminal domain (amino acids 382–670; ATF6-C) by co-IP assay of transfected HEK293T cells. HA, haemagglutinin; WT, wild type. Tubulin was used as a loading control in **c** and **d**.

¹Clayton Foundation Laboratories for Peptide Biology, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, California 92037, USA.

gene expression^{4–9}. CRTC2 was found to interact with the transcriptionally active cytoplasmic amino-terminal domain (ATF6-N; amino acids 1–381) but not with the ER luminal carboxy-terminal domain of ATF6 α (ATF6-C; amino acids 382–670) (Fig. 1d). Conversely, ATF6 α associated with an N-terminal CRTC2 polypeptide (amino acids 1–120) that also mediates an interaction with CREB (Supplementary Fig. 2b)^{15,16}.

We examined the effects of ER stress on the recruitment of CRTC2 to ATF6 α -regulated genes. Under basal conditions, about one-third of cellular CRTC2 was localized to the cytoplasmic surface of the ER (Supplementary Fig. 3). After exposure of primary hepatocytes to THA or TUN, ATF6 α and CRTC2 shuttled to the nucleus where they occupied the *Xbp1* promoter⁷ (Fig. 2a, top). ATF6 α overexpression augmented CRTC2 occupancy, whereas RNAi-mediated knockdown of *Atf6a* blocked it (Fig. 2a, bottom). Furthermore, CRTC2

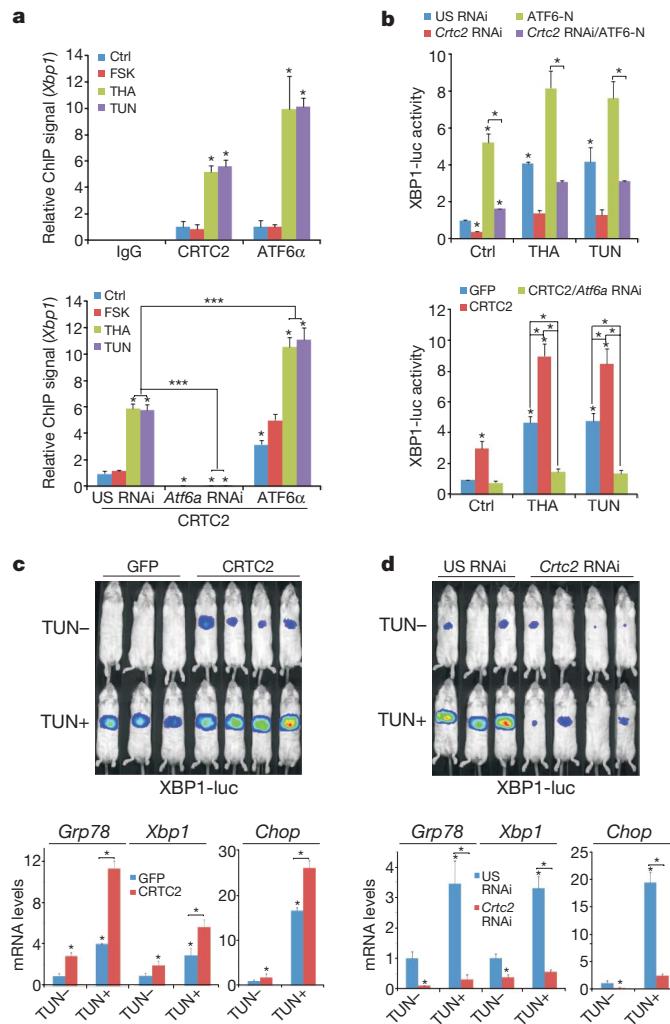


Figure 2 | CRTC2 stimulates the expression of ER quality control genes through an association with ATF6 α . **a**, Top, ChIP assay showing occupancy of CRTC2 and ATF6 α over the *Xbp1* promoter in primary hepatocytes exposed to THA or TUN. Bottom, effect of adenoviral ATF6 α overexpression or RNAi-mediated knockdown (*Atf6a* RNAi) on CRTC2 occupancy. *P < 0.01, ***P < 0.001, n = 3. Ctrl, control. **b**, Top, effect of adenoviral ATF6-N on Ad-XBP1-luc reporter activity in control (unspecific (US) RNAi) and CRTC2-depleted (*Crtc2* RNAi) cells. Bottom, effect of Ad-CRTC2 on Ad-XBP1-luc reporter activity in control and ATF6 α -depleted hepatocytes. *P < 0.01, n = 4. **c**, **d**, Top, hepatic Ad-XBP1-luc reporter activity in mice injected intraperitoneally with TUN or vehicle. GFP, green fluorescent protein. Shown are the effects of Ad-CRTC2 overexpression (**c**) or Ad-*Crtc2* RNAi (**d**) on Ad-XBP1-luc activity (top) and on hepatic mRNA amounts for ATF6 α -regulated genes (bottom). *P < 0.01, n = 4. All error bars are s.e.m.

overexpression increased Ad-XBP1-luc reporter activity, whereas RNAi-mediated depletion of *Crtc2* reduced it (Fig. 2b). Consistent with a requirement for ATF6 α , CRTC2 did not upregulate Ad-XBP1-luc activity when cells were depleted of ATF6 α .

During *ad libitum* feeding, CRTC2 activity is silenced by phosphorylation at Ser 171 by salt inducible kinase 2 (SIK2); these effects are reversed during fasting, when glucagon inhibits SIK2 activity via protein-kinase-A-mediated phosphorylation¹. Pointing to a role for this kinase in the ER stress response, SIK2 overexpression disrupted Ad-XBP1-luc reporter activity in ER-stressed hepatocytes, whereas RNAi-mediated knockdown augmented it (Supplementary Fig. 4a, b). Conversely, overexpression of a phosphorylation-defective, active Ser171Ala-mutant CRTC2 increased Ad-XBP1-luc reporter activity constitutively in cells co-expressing active ATF6-N (Supplementary Fig. 4c).

We examined whether CRTC2 modulates hepatic ER stress responses *in vivo*. Triggering of the ER stress pathway by intraperitoneal injection of TUN increased Ad-XBP1-luc reporter activity and ER stress gene expression in both fasted and *ad libitum*-fed mice (Fig. 2c; *Xbp1*, *Grp78* (also known as *Hspa5*) and *Chop* (*Ddit3*); Supplementary Figs 4–6); these effects were attenuated by RNAi-mediated depletion of hepatic *Atf6a* (Supplementary Fig. 7). Similar to its effects in hepatocytes, Ad-CRTC2 expression also enhanced hepatic Ad-XBP1-luc reporter activity and ER stress gene expression in TUN-injected mice, whereas RNAi-mediated depletion of CRTC2 downregulated it (Fig. 2c, d and Supplementary Fig. 8). Taken together, these results indicate that CRTC2 promotes the expression of ER quality control genes in liver by an association with ATF6 α .

We proposed that ATF6 α could interfere with the induction of the gluconeogenic program through the CREB–CRTC2 pathway if cellular levels of CRTC2 were limiting. Supporting this idea, exposure of primary hepatocytes to THA or TUN increased the binding of ATF6 α (p50, the active ATF6 α fragment) to CRTC2, and reciprocally reduced the amounts of CRTC2 associated with CREB (Fig. 3a, top). Overexpression of nuclear, active ATF6-N also decreased the CREB–CRTC2 interaction, and blocked the recruitment of CRTC2 to the gluconeogenic G6Pase (also known as G6pc) promoter in FSK-treated cells (Fig. 3a and Supplementary Fig. 9). ATF6-N expression also decreased Ad-CRE-luc reporter activity, gluconeogenic gene expression, and glucose output from primary hepatocytes, whereas RNAi-mediated depletion of *Atf6a* increased it (Fig. 3b and Supplementary Fig. 9). Confirming the importance of the ATF6 α –CRTC2 interaction, a mutant (Arg337Ala) ATF6 α polypeptide with lower affinity for CRTC2 was less potent in disrupting CREB activity relative to the wild-type (Supplementary Fig. 10).

Because ATF6 α and CREB bind to the same domain in CRTC2, CREB may reciprocally downregulate ATF6 α activity. Indeed, Ad-CREB expression not only reduced Ad-XBP1-luc reporter activity, but also increased CRE-luc reporter activity and glucose production in hepatocytes expressing ATF6-N (Fig. 3b and Supplementary Fig. 11). Taken together, these results indicate that CREB and ATF6 α exert counter-regulatory effects on gluconeogenesis, in part by competing for CRTC2.

We tested the role of ATF6 α in modulating the glucose balance *in vivo*. Modest hepatic overexpression of ATF6 α lowered the fasting gluconeogenic profile—which consists of hepatic Ad-CRE-luc activity, gluconeogenic gene expression, and circulating blood glucose concentrations—in control mice, and to a greater extent in mice injected intraperitoneally with TUN (Fig. 3c and Supplementary Fig. 12a). In contrast, RNAi-mediated depletion of hepatic *Atf6a* increased the fasting gluconeogenic profile in both control and TUN-injected animals (Fig. 3d and Supplementary Fig. 12b).

Considering that ER stress is chronically upregulated in obesity³ and that ATF6 α undergoes proteasome-dependent degradation when ER stress is prolonged⁸, we tested whether hepatic ATF6 α activity is altered in this setting. Relative to lean controls, obese

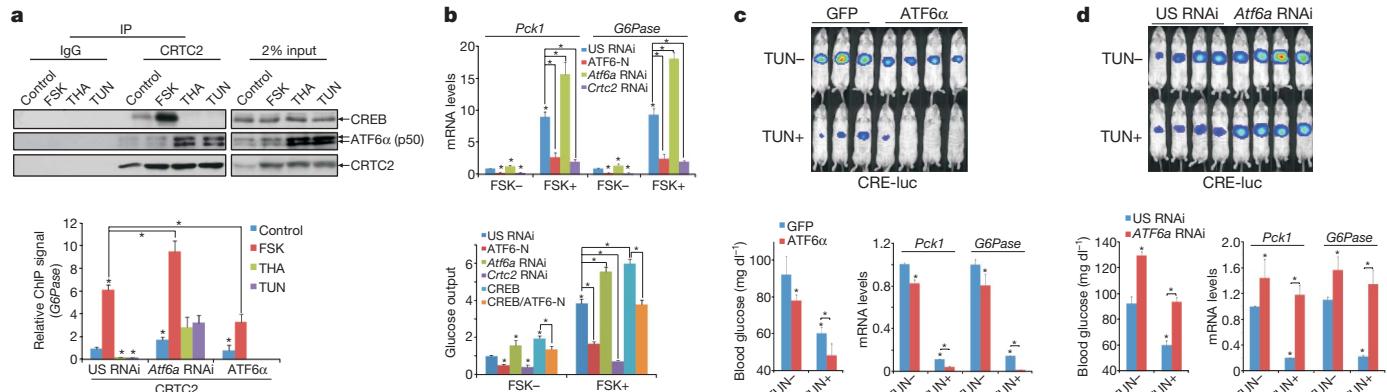


Figure 3 | CRTC2 mediates cross-talk between hepatic ER stress and fasting pathways. **a**, Top, immunoblot showing recovery of CREB from co-immunoprecipitations of CRTC2 prepared from nuclear extracts of primary hepatocytes exposed to FSK, TUN or THA. Bottom, effect of ATF6 α overexpression or RNAi-mediated depletion (*Atf6a* RNAi) on CRTC2 occupancy over the *G6Pase* promoter. *P < 0.01, n = 4. **b**, Effect of ATF6 α overexpression or RNAi-mediated knockdown on gluconeogenic gene expression (top; n = 4) and glucose secretion (bottom; n = 5) in primary hepatocytes exposed to FSK. *P < 0.01. **c, d**, Effect of ATF6 α overexpression (c) or RNAi-mediated depletion (d) on hepatic Ad-CRE-luc activity (top), circulating blood glucose concentrations (bottom left), and gluconeogenic gene expression in fasted mice injected intraperitoneally with TUN or vehicle. *P < 0.01, n = 4. All error bars are s.e.m.

(*ob/ob*, *db/db*) mice showed lower Ad-XBP1-luc reporter activities, and they had reduced amounts of hepatic ATF6 α protein (Fig. 4a and Supplementary Fig. 13). In contrast, Ad-CRE-luc activity, gluconeogenic gene expression, and circulating blood glucose concentrations were all increased in the *ob/ob* and *db/db* animals.

Because RNAi-mediated depletion of *Atf6a* increases CRE-luc activity and hepatic glucose output, we wondered whether the obesity-related loss of hepatic ATF6 α would have similar effects on CREB and CRTC2. In that case, ATF6 α overexpression should improve glucose balance by lowering gluconeogenic gene expression. Supporting this notion, adenoviral expression of active ATF6-N reduced fasting

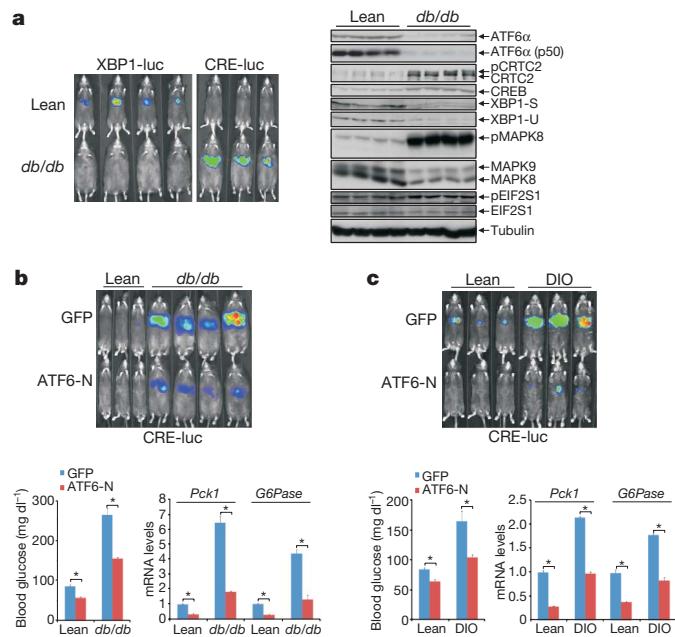


Figure 4 | Reciprocal downregulation of ATF6 α and upregulation of CREB in obesity. **a**, Left, hepatic Ad-XBP1-luc and Ad-CRE-luc activities in obese *db/db* mice relative to lean controls. Right, immunoblot of hepatic extracts from wild-type (lean) and *db/db* mice showing protein amounts of ATF6 α and other ER stress markers. **b, c**, Effect of Ad-ATF6-N relative to Ad-GFP control on hepatic Ad-CRE-luc activity (top), blood glucose concentrations (bottom left), and gluconeogenic gene expression (bottom right) in *db/db* (b) and high-fat diet-fed (DIO, c) mice compared to lean controls. *P < 0.01, n = 4. All error bars are s.e.m.

expression (top; n = 4) and glucose secretion (bottom; n = 5) in primary hepatocytes exposed to FSK. *P < 0.01. **c, d**, Effect of ATF6 α overexpression (c) or RNAi-mediated depletion (d) on hepatic Ad-CRE-luc activity (top), circulating blood glucose concentrations (bottom left), and gluconeogenic gene expression in fasted mice injected intraperitoneally with TUN or vehicle. *P < 0.01, n = 4. All error bars are s.e.m.

hepatic Ad-CRE-luc activity as well as messenger RNA amounts for gluconeogenic genes in both *db/db* and high-fat diet-fed mice (Fig. 4b, c and Supplementary Fig. 14). Overexpression of ATF6-N also lowered circulating blood glucose levels and enhanced glucose tolerance (Fig. 4b, c and Supplementary Fig. 15). Arguing against an important effect of ATF6-N on insulin signalling per se, hepatic amounts for inactive phosphorylated (Ser 307) insulin receptor substrate 1 (IRS1) or active phosphorylated (Thr 308) AKT appeared comparable between ATF6-N-expressing and control mice (Supplementary Fig. 15).

Taken together, these results indicate that CRTC2 functions as a dual sensor for fasting and ER stress signals (Supplementary Fig. 16). The attendant cross-talk between these pathways seems to protect against excessive increases in hepatic gluconeogenesis that otherwise lead to chronic hyperglycemia in obesity. In addition to its effects on ATF6 α , chronic ER stress has also been found to increase hepatic gluconeogenesis and lipogenesis by triggering other branches of the unfolded protein response¹⁷. Future studies should reveal the extent to which these pathways promote glucose intolerance through CREB or other bZIP transcription factors.

METHODS SUMMARY

Adenoviruses were delivered to 8–10-week-old male mice by tail-vein injection. Mice expressing CRE-luc or XBP1-luc reporters were imaged on an IVIS 100 Imaging System under *ad libitum* feeding conditions or after fasting¹. Hepatic ER stress was induced *in vivo* by intraperitoneal injection of TUN⁶; effects of TUN on reporter activity and hepatic gene expression were evaluated 10 h after TUN administration. For glucose tolerance tests, mice were fasted overnight and then injected intraperitoneally with glucose. The effects of ER stress on glucose output, luciferase reporter activity, and gluconeogenic gene expression were also examined using primary hepatocyte cultures¹². The relative occupancy of ATF6, CREB and TORC2 over relevant promoters in hepatocytes was determined by chromatin immunoprecipitation (ChIP) assay². CRTC2-associated proteins were identified using immunoprecipitates of endogenous and epitope-tagged CRTC2 prepared from HEK293T cells¹², and analysed by electrospray ionization tandem mass spectrometry.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 17 March; accepted 7 May 2009.

Published online 21 June 2009.

- Dentin, R. et al. Insulin modulates gluconeogenesis by inhibition of the coactivator TORC2. *Nature* **449**, 366–369 (2007).
- Koo, S. H. et al. The CREB coactivator TORC2 is a key regulator of fasting glucose metabolism. *Nature* **437**, 1109–1111 (2005).
- Ozcan, U. et al. Endoplasmic reticulum stress links obesity, insulin action, and type 2 diabetes. *Science* **306**, 457–461 (2004).

4. Haze, K., Yoshida, H., Yanagi, H., Yura, T. & Mori, K. Mammalian transcription factor ATF6 is synthesized as a transmembrane protein and activated by proteolysis in response to endoplasmic reticulum stress. *Mol. Biol. Cell* **10**, 3787–3799 (1999).
5. Adachi, Y. *et al.* ATF6 is a transcription factor specializing in the regulation of quality control proteins in the endoplasmic reticulum. *Cell Struct. Funct.* **33**, 75–89 (2008).
6. Wu, J. *et al.* ATF6 α optimizes long-term endoplasmic reticulum function to protect cells from chronic stress. *Dev. Cell* **13**, 351–364 (2007).
7. Yoshida, H., Matsui, T., Yamamoto, A., Okada, T. & Mori, K. XBP1 mRNA is induced by ATF6 and spliced by IRE1 in response to ER stress to produce a highly active transcription factor. *Cell* **107**, 881–891 (2001).
8. Lin, J. H. *et al.* IRE1 signaling affects cell fate during the unfolded protein response. *Science* **318**, 944–949 (2007).
9. Yamamoto, K. *et al.* Transcriptional induction of mammalian ER quality control proteins is mediated by single or combined action of ATF6 α and XBP1. *Dev. Cell* **13**, 365–376 (2007).
10. Kahn, B. B. & Flier, J. S. Obesity and insulin resistance. *J. Clin. Invest.* **106**, 473–481 (2000).
11. Saltiel, A. R. New perspectives into the molecular pathogenesis and treatment of type 2 diabetes. *Cell* **104**, 517–529 (2001).
12. Screamton, R. A. *et al.* The CREB coactivator TORC2 functions as a calcium- and cAMP-sensitive coincidence detector. *Cell* **119**, 61–74 (2004).
13. Bittinger, M. A. *et al.* Activation of cAMP response element-mediated gene expression by regulated nuclear transport of TORC proteins. *Curr. Biol.* **14**, 2156–2161 (2004).
14. Yoshida, H. *et al.* ATF6 activated by proteolysis binds in the presence of NF-Y (CBF) directly to the *cis*-acting element responsible for the mammalian unfolded protein response. *Mol. Cell. Biol.* **20**, 6755–6767 (2000).
15. Iourgenko, V. *et al.* Identification of a family of cAMP response element-binding protein coactivators by genome-scale functional analysis in mammalian cells. *Proc. Natl Acad. Sci. USA* **100**, 12147–12152 (2003).
16. Conkright, M. D. *et al.* TORCs: transducers of regulated CREB activity. *Mol. Cell* **12**, 413–423 (2003).
17. Oyadomari, S., Harding, H. P., Zhang, Y., Oyadomari, M. & Ron, D. Dephosphorylation of translation initiation factor 2 α enhances glucose tolerance and attenuates hepatosteatosis in mice. *Cell Metab.* **7**, 520–532 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by grants from the National Institutes of Health, by the Clayton Foundation for Medical Research, by the Keckhefer Foundation, and by the Vincent J. Coates Foundation. We thank N. Miller, S. Hedrick, and Y. Liu for technical assistance and helpful discussions.

Author Contributions Y.W. and L.V. performed *in vivo* imaging studies; Y.W. performed *in vitro* and biochemical studies; W.H.F. carried out mass spectrometry analysis. Y.W. and M.M. designed the study, analysed the data, and wrote the paper. All authors reviewed and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.M. (montminy@salk.edu).

METHODS

Mouse strains and adenovirus. Adenoviruses (1×10^8 plaque forming units (p.f.u.) *Crtc2* RNAi, CRTC2, *Atf6a* RNAi, ATF6 α , ATF6-N (amino acids 1–381), or unspecific RNAi, 10^9 p.f.u. CRE-luc or XBP1-luc reporter, 5×10^7 p.f.u. RSV- β -gal (Rous sarcoma virus promoter)) were delivered to 8–10-week-old male C57BL/6J-Tyrc-2J/J, C57BL/6J, B6.V-lep<ob>/J, B6.Cg-m+/+Lepr<db>/J or diet-induced obesity mice by tail-vein injection. These mice were purchased from Jackson Laboratories. All mice were adapted to their environment for 1 week before study, and were housed in colony cages with 12 h light/dark cycle in a temperature-controlled environment. For *in vivo* imaging experiments, mice were imaged on days 3–5 after adenovirus delivery. Wild-type CRTC2, mutant Ser171Ala CRTC2, *Crtc2* RNAi, unspecific RNAi, CRE-luc and RSV β -gal adenoviruses have been described previously¹. Ad-*Atf6a* RNAi was constructed using the sequence 5'-GGGAGTCAGACCTATGGAGCCC-3'. XBP1-luc was made from the mouse *Xbp1* promoter region spanning from -614 to -1.

In vivo imaging. Mice were imaged as described¹ under *ad libitum* feeding conditions or after fasting for 6 h. Mice were injected intraperitoneally with glucagon (100 mg kg⁻¹; Sigma), TUN (1 g kg⁻¹, Calbiochem)⁶, or vehicle. Before imaging, mice were injected intraperitoneally with 50 mg kg⁻¹ Nembutal (Abbott Laboratories) and 100 mg kg⁻¹ sterile firefly D-luciferin (Biosynth AG). Mice were imaged on the IVIS 100 Imaging System, and analysed with Living Image software (Xenogen) 1 h after glucagon injection, and 10 h after TUN administration.

In vivo analysis. Mouse tissues were sonicated, centrifuged and supernatants were reserved for β -gal activity, protein determinations, and SDS-PAGE analysis. Blood glucose values were determined using a LifeScan automatic glucometer. Glucose tolerance tests were performed by glucose intraperitoneal administration (1 g kg⁻¹) after overnight fasting on day 5 after injection with Ad-GFP or Ad-ATF6-N adenovirus.

Cell culture and fractionation. Mouse CRL-2189 (ATCC) cells were cultured in DMEM. Mouse primary hepatocytes were isolated and cultured as previously described¹. For reporter studies, Ad-CRE-luc or Ad-XBP1-luc infected hepatocytes (1 p.f.u. per cell) were exposed to FSK (10 μ m) for 4 h, or THA (5 μ g ml⁻¹) or TUN (200 nM) for 12–15 h. For cyclosporine A (5 μ M) inhibition, hepatocytes were pre-treated with cyclosporine A for 1–2 h. For double stimulation with

FSK and THA, or with FSK and TUN, hepatocytes were pre-treated with THA or TUN for 4 h, and then co-incubated with FSK for another 4 h. Glucose output from primary hepatocytes was determined enzymatically, after 1 h collection in glucose-free M199 media supplemented with 10 mM lactate and 1 mM pyruvate¹². Cellular fractionation was performed as reported¹⁸.

Immunoblot, immunoprecipitation, GST pull-down and immunostaining. Immunoblot, immunoprecipitation, glutathione S-transferase (GST) pull-down and immunostaining assays were performed as described¹². CRTC2, pCREB and CREB antibodies were previously described². The antibody anti-ATF6 α was purchased from IMGENEX, anti-tubulin from Upstate, anti-haemagglutinin (HA) from Covance, anti-EEA1 from BD Pharmingen, anti-KDEL, GRP94 and TGN46 from Abcam, anti-XBP1, JNK and cytochrome C from Santa Cruz, and anti-pJNK, pIF2 α , eIF2 α , pAKT, AKT, pIRS1 and IRS1 from Cell Signaling.

Chromatin immunoprecipitation and quantitative PCR. ChIP assays were performed as previously described². The primers for *G6Pase* (5'-GGAGGGCA GCCTCTAGCACTGTCAA-3', 5'-TCAGTCTGTAGGTCAATCCAGCCCT-3') and *Xbp1* (5'-GGCCCCAGTTGCACGGCGGAGAACAA-3', 5'-CACCGCCCC GTGGCCTCCTGCCGC-3') were used for ChIP analysis. All signals were normalized to the input chromatin signals. Total cellular RNAs from whole liver or from primary hepatocytes were extracted using the RNeasy kit (Qiagen). mRNA levels were measured as previously described¹.

Mass spectrometry. Immunoprecipitates of endogenous and HA-tagged CRTC2 from HEK293T were prepared for mass spectrometric studies as previously reported¹², and analysed by electrospray ionization tandem mass spectrometry on a Thermo LTQ Orbitrap instrument.

Statistical analyses. All studies were performed on at least three independent occasions. Results are reported as mean and s.e.m. The comparison of different groups was carried out using a two-tailed unpaired Student's *t*-test. Differences were considered statistically significant at $P < 0.05$.

18. Robinson, L. J., Pang, S., Harris, D. S., Heuser, J. & James, D. E. Translocation of the glucose transporter (GLUT4) to the cell surface in permeabilized 3T3-L1 adipocytes: effects of ATP insulin, and GTP gamma S and localization of GLUT4 to clathrin lattices. *J. Cell Biol.* 117, 1181–1196 (1992).

CORRIGENDUM

doi:10.1038/nature08232

**Life without a wall or division machine in
*Bacillus subtilis***

M. Leaver, P. Domínguez-Cuevas, J. M. Coxhead, R. A. Daniel
& J. Errington

Nature 457, 849–853 (2009)

It has been drawn to our attention that proliferation by a ‘budding’ process, similar to the one we described, has been reported many years ago for pleuropneumonia-like organisms and L-form bacteria^{1,2}. The implication that this mode of replication is common to a wide range of bacteria supports the idea that ‘extrusion-resolution’ is representative of an ancient mode of proliferation invented before the divergence of the main groups of modern bacteria.

1. Kandler, G. & Kandler, O. Untersuchungen über die Morphologie und die Vermehrung der pleuropneumonie-ähnlichen Organismen und der L-Phase der Bakterien. I. Lichtmikroskopische Untersuchungen. *Arch. Mikrobiol.* 21, 178–201 (1954).
2. Kandler, G., Kandler, O. & Huber, O. Untersuchungen über die Morphologie und die Vermehrung der pleuropneumonieähnlichen Organismen und der L-Phase der Bakterien. II. Elektronenmikroskopische Untersuchungen. *Arch. Mikrobiol.* 21, 202–216 (1954).

Big opportunities in a small world

Nanomedicine has started to gather momentum in recent years, and cutting-edge jobs abound for those with the right training. **Virginia Gewin** takes a closer look.

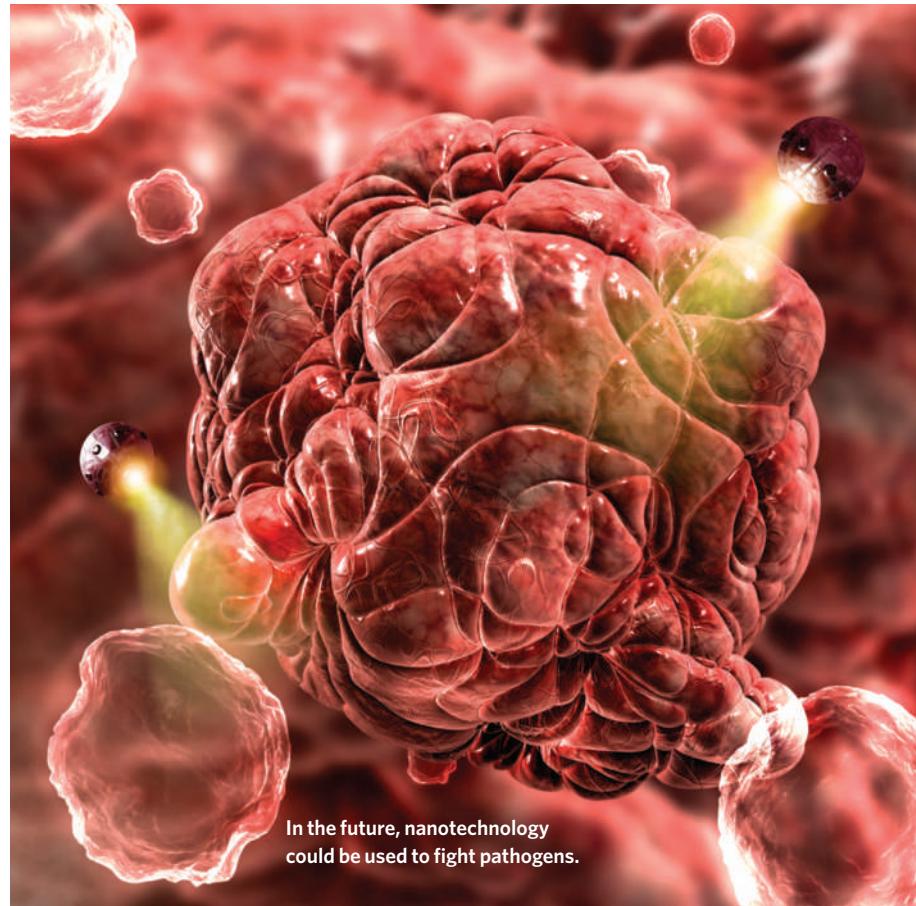
Mauro Ferrari is on a mission to make Houston a major hub for nanomedicine. Housed at the University of Texas Health Science Center, Ferrari's lab will become the Department of Nanomedicine and Biomedical Engineering in September. According to Ferrari, it will be the first nanomedicine department at a US medical school. The nanomedicine professor, whose team is developing nanosized diagnostic devices to treat cancer and cardiovascular disease, is in an enviable position, especially given the current economic climate. Ferrari plans to recruit another 30 researchers to complete his 100-member academic research team, and he has co-founded two companies. NanoMedical Systems, in Austin, Texas, is developing a nanomaterial-based drug-delivery system, and Leonardo Biosystems in Houston is researching nanotechnology-based cancer therapeutics.

Houston is already well known for its prowess in the nanoscopic field. The city was home to the 1985 discovery of spherical carbon-based fullerenes known as 'buckyballs'. That work was later awarded the 1996 Nobel Prize in Chemistry. Yet so far, nanotech advances have been more incremental than monumental. This could be set to change as research funds start to flow, nanomedicines head to clinical trials and entrepreneurial academics aim to incorporate nanomedicine into mainstream medical care.

Governments around the world are banking on high economic returns as they invest in a field that aims to use atomic- or molecular-level techniques to repair damaged tissue or diagnose, treat or prevent disease. Academics are forging multidisciplinary teams of scientists, engineers and clinicians eager to test nanosolutions to medical problems. But there are challenges. The field needs greater numbers of highly trained students, and a sound regulatory infrastructure.

Entrepreneurial academics

Houston isn't the only budding nanomedicine hub. In the United Kingdom, Swansea University, with input from Ferrari, has brought together teams of clinicians, life scientists, engineers, physical scientists and industry professionals. The university will break ground this summer on its £21.6-million (US\$35-million) Centre for NanoHealth, which will be the cornerstone



In the future, nanotechnology could be used to fight pathogens.

of a plan to grow its schools of medicine and engineering. The building is slated to open in 2010, and Swansea University administrators are searching for researchers eager to cross disciplinary boundaries. Over the next year, they will begin to fill 12 core research posts, including the centre manager, academic posts and clinicians to support the centre's mission.

The key to successful growth of a nanomedicine hub, say the organizers, is forging industry partnerships. One of Swansea University's spin-off companies — Haemair, also based in Swansea — has developed a blood oxygenating system to help patients with impaired breathing.

The centre's co-director Steven Conlan says that four more companies are being developed. Based on past job-creation initiatives, Gareth Morgan, the head of Swansea's school of medicine, expects the centre to create up to 450 jobs over the next

five years, many in small and medium-sized partner companies. Conlan says the centre will build on the university's strengths in engineering and physical sciences to first focus on developing nanotech devices, such as *in vivo* sensors. People trained in rheology, scanning microscopy, biochemistry, device fabrication and chemistry will be in demand as the centre develops high-tech platforms to be licensed and commercialized.

"We need people who have developed an intuition about multiscale systems."

— Randy Goodall

Such institutional configurations mirror a merging of nanotechnology and medical career paths. Shad Thaxton was halfway through his urology training in 2002 when he read about the nanotechnology research in Chad Mirkin's laboratory. Both were at Northwestern University in Illinois — Mirkin at the Evanston campus and Thaxton in Chicago — and they began exploring nanotechnology's potential impact on medicine. Thaxton completed his medical degree and built on his

clinical and research experience to learn how to synthesize and characterize nanoparticles. With an eye towards repairing atherosclerotic damage, Thaxton — now an assistant professor at Northwestern University — is working to develop a synthetic nanoparticle that mimics the action of high-density lipoprotein, which is thought to help remove cholesterol from within arteries.

Mirkin, a chemist by training, has so far integrated three medical doctors into his Northwestern nanomedicine research group. "The doctors give us the opportunity not just to do great science, but to explore nanotherapeutics as well," he says. With more than \$450 million in funding, Mirkin's International Institute for Nanotechnology at

Northwestern University covers a broad range of topics — from intracellular gene regulation to developing new materials that can penetrate tumours.

Always hiring

Like Ferrari, Mirkin has an entrepreneurial bent. He formed Nanosphere, a molecular-diagnostics company in Northbrook, Illinois. Mirkin typically recruits individuals with strong backgrounds in physics, chemistry or medicine who also demonstrate the ability to conduct novel, interdisciplinary nanotechnology research. "We have one of the largest operations in the world, and we're always hiring," he says.

Others are boasting similar good fortune. Freddy Boey Yin Chiang, head of materials science and engineering at Nanyang Technological University in Singapore, has benefited from Singapore's growing financial support for nanomedicine. He has recruited graduate students, postdocs and researchers from the United States and Europe, and he hopes to lure more as he expands his lab by 20 faculty positions during the next few years. "My problem is not capacity, it is finding good people," he says.

Many labs share Boey's dilemma. Although there is a multitude of opportunities, finding people with the requisite skills for this interdisciplinary field can be difficult. "We need people who have developed an intuition about multiscale systems — those who can think about the particle or structure that is one billionth of a metre and its potential role in a device that is a few centimetres in size," says Randy Goodall, chief executive of NanoMedical Systems.

Vladimir Zharov, director of the Phillips Classic Laser and Nanomedicine



Mauro Ferrari, left, and Shad Thaxton both work at the cutting edge of nanomedicine.

Laboratories at the University of Arkansas for Medical Sciences in Little Rock, has had difficulty finding individuals with the mix of experience necessary to develop laser-activated gold nanoparticles to treat cancer. In May he received US National Cancer Institute funding to conduct a first-of-its-kind clinical trial of nanomedicine technology related to early cancer diagnosis.

Zharov sees a large education gap that, at the moment, 'on-the-job' training must bridge. "We need more combined degrees," he says. Unfortunately, a lack of qualified instructors has meant that such cross-disciplinary training is not easy to accomplish. "Training may be the weakest piece of the nanomedicine enterprise," agrees Laurent Lévy, chief executive of Nanobiotix, a nanomedicine company in Paris working to achieve more potent, targeted radiotherapy.

Medical schools are beginning to heed the call. Starting this autumn, University College London (UCL) will offer a master's of science in nanotechnology and regenerative medicine. "It takes a long time to get into the nanotech field, so we thought it would be useful to have a course that would help students to start a research project and lead them to a PhD," says Alexander Seifalian, a biomaterials researcher and programme course leader. He says the course includes a hands-on workshop to expose students to nanotech research opportunities related to regenerative medicine. UCL will accept 10 students this year, with plans to increase enrolment to 25 students in subsequent years.

In addition to its doctorate programmes, Swansea University will start running an

"Training may be the weakest piece of the nanomedicine enterprise."

— Laurent Lévy

MS in nanomedicine this year. "We need a new breed of students," says Conlan. The university is also looking to build industry-academia partnerships that would provide industry experience for postdocs.

Regulatory landscape

As nanomedicine technologies inch closer to market, establishing a regulatory road map has the potential to create more career opportunities. To date, most researchers have focused on developing biomaterials or designing better particles. But regulators with science training will be in demand to assess safety and efficacy. The need to understand the health effects of nanotech exposure is growing with the field, says Sally Tinkle, senior science adviser at the US National Institute of Environmental Health Sciences. In April, the institute earmarked some of its federal economic stimulus funds to support nanomaterials and medicinal research, hoping to develop tools and methods to assess the safety of nanomaterials. "When you measure a drug-delivery system that targets a tumour, you also need to measure clearance of nanomaterials from the system," says Tinkle.

Likewise, preclinical characterization of nanoparticles intended for cancer therapeutics and diagnostics is in demand. Scott McNeil runs the Nanotechnology Characterization Laboratory, a National Cancer Institute facility in Frederick, Maryland. He started the programme in 2005 with eight scientists. Since then, the lab has helped characterize, at some level, 150 different nanoparticle types. Now, his current team of 20 chemists, physicists, biologists and toxicologists are all kept busy conducting full characterizations, from particle-size distribution to *in vivo* toxicity, of new nanoparticle submissions.

McNeil says that scientists working with the first generation of nanomedicinal particles simply took advantage of the small particle size to achieve novel drug properties — for example, the reformulation of the chemotherapy drug Doxil (doxorubicin) to boost the amount of chemotherapeutic drug routed to the malignant cells. Now, says McNeil, researchers have learned how to engineer specific properties at the nanoscale. This requires additional characterization work by his team. He is recruiting for senior researchers and postdocs to keep up with demand.

Like work at most frontiers, nanomedicine is rife with opportunities and challenges. "Nanomedicine has unbelievable transformational potential," says Ferrari. "But we must be as brave, persistent and smart as the biotechnology innovators were when that field was emerging 20 years ago."

Virginia Gewin is a science writer based in Portland, Oregon.

The problem of Junior

Parent trap.

Swapna Kishore

Positronic circuits, decision algorithms, laws of robotics, fuzzy logic — it wasn't just difficult, it was boring. Roy would have rather experimented with the recipe in *Gourmet Fare* (Clara loved fine dining) but the note from Junior's school (yet again) meant that Roy could not ignore the problem of Junior any more.

Junior chattered away, unrepentant, a bouncy cheer on his chubby face, so scrubbed-clean, an ideal seven year old. Roy kept glancing at him as he drove him back from school. How could Junior just dismiss the scowling principal? Or ignore nanny's head rolling on the floor, its lips muttering, "Good boys don't tell lies."

"Why did you do it?" Roy asked, unable to stop himself.

"Do what?" Junior looked mildly puzzled. "Oh, that stupid nanny. Old version. Head came off so easily, too. The school should get upgrades."

"Your nanny wasn't a clunky metal model, or even an icky-plastic one." That would have been okay. "Nanny was a positronic humanoid." Roy almost added, like you, but Junior didn't know, so he took a deep breath and swallowed the words. "Didn't it hurt you to yank her head off?" Junior's circuits should have stopped him from violating the third law.

"Forget it," Junior said. "Say, will Mom get my PlayStation today?"

Clara. Roy's guts twisted. She would be so unhappy about this.

Six months ago, Clara said: "I want a son."

Beethoven background, her favourite symphony. Dark coffee at her elbow, exactly as she liked it. New chairs, the right softness, the right peach shade. Another perfect evening arranged by Roy. So why this?

"Am I not good enough?" he asked, hurt.

"We'll be a complete family."

"But you'll need surgery."

Clara laughed. "Don't be funny. Constructs are offering discounts if we try their new version."

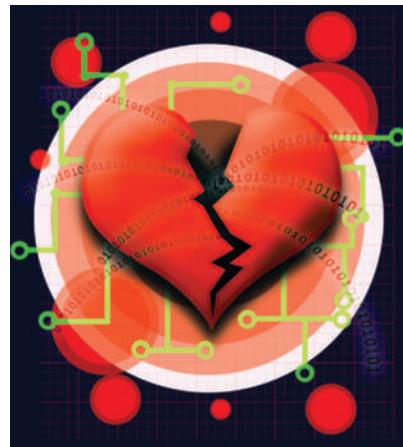
She chose the age (seven, so toilet-trained, distinct personality), gender (boy, no cutie girl), intelligence (high, of course) and docility (low, because a boy must be a boy, or what's the point?). Roy soaked in her smile, and felt his heart warm up in response.

For past 'family fun-time memories', Clara selected parameters for beach holidays and picnics and auto-generated photo albums. When Roy suggested watching the

coded-in past, she laughed. "Seven years' worth? Don't be silly." She collapsed salient memories to a three-day capsule, gave the bot a hazy amnesia for the rest.

Junior arrived.

Clara gave up the pick-him-up-from-school in a week, the let's-do-homework-together in two. She soon tired of Junior's prattle about homework, class bullies and maths lessons. Within a month of Junior's



delivery, she shrugged off Roy's updates (and complaints) on their 'son', and reminded him that she needed enough me-time to recharge herself; her job had responsibilities that needed full attention.

Roy suggested 'returning' Junior to Constructs, but Clara sniggered that surely a househusband could handle a small bot? It almost hurt, but he loved her too much to mind for long.

He bought a course on fathering bots.

The law was clear on the nanny episode: boticide by a bot was a severity-3 deviation; Junior should be reprogrammed or destructed.

Problem was, Clara still loved the concept of Junior. His behavioural lapses were, to her, a failure of her specifications, an embarrassment for a top-notch professional. It was as if, by pretending all was well, they would be the ideal happy family.

Roy squirmed. She shouldn't have requisitioned a high anti-docility.

But he loved her and he wanted her to be happy.

"Junior destroyed a humanoid nanny," he told Clara that evening.

She paled.

"I'm sorry," he said. "I tried my best."

"No, no, not your fault." She flipped on her PDA, entered some codes. "This is awful; I

hate it." She got up and began pacing.

That's when Junior walked in, grinning and muddy, holding a soccer ball. Like nothing had happened.

And Roy got a brainwave: he may have failed as father, but he could save Clara some agony now.

If only Junior didn't look so real.

Destructing him — it — was legally allowed — no, required — because a bot had destroyed another.

At times like this, when Roy forced himself to do something he hated because of Clara, he suspected this was not what love should be. Then he thought of her smile.

Clara's PDA lay on the table, with Junior's destruct code on top.

Damn.

Roy fought his nausea and clicked 'Confirm.'

Junior froze, and then slumped, limbs tangled in impossible ways. The ball slipped out of his hands and thumped to the floor. A faint smell of burnt circuitry began spreading.

"Roy?" Clara looked at the awkward arrangement of dismantled body parts, and then at Roy. "Why did you...?"

Roy wanted to throw up, but he kept his face expressionless; he didn't want to alarm Clara. "Junior was dangerous, but you would have hated destroying him."

Clara blinked rapidly. Her mouth fell open; she closed it.

"Next time, specify less naughtiness, okay?" he said softly.

Tenderness suffused her face. "I really messed up the specifications, huh? So many contradictions."

He blew her a kiss. His head hurt. "It's okay."

"Was it tough destructing Junior?" Clara squeezed his hand. Her gentle, almost wistful smile made the migraine worth it.

"No," he lied.

Pause.

"I love you," he added. That, at least, was true.

"Too much, I think." She sighed. "I loved you."

She hugged him and stepped back. Something glittered in her hand. Her eyes were moist.

"But your circuits didn't work either," she whispered, and everything exploded inside him.

Swapna Kishore is a software consultant in Bangalore, India. She has published technical books and a business novel. She also writes speculative fiction.